UNIVERSITY OF MICHIGAN
SEP 12 1961
BUSINESS ADMINISTRATION LIBRARY

Journal of Advertising Research

Vol. 1, No. 5 SEPTEMBER 1961

Treatment of the Non-Response Problem
TORE DALENIUS 1

A Comparison of Eight Audience Estimates
J. STEVENS STOCK 9

A Survey of Case History Advertising
GEORGE FISK 16

Selecting Commercial Spokesmen
ARTHUR J. KOVER AND SEYMOUR LIEBERMAN 22

FEDERAL STATISTICS IN ADVERTISING 26

ON METHODS 28

RESEARCH IN REVIEW 33

PUBLICATIONS RECEIVED 39

EDITORIAL 40

Published by the Advertising Research Foundation

The Journal of Advertising Research

is published quarterly by the Advertising Research Foundation and sent free to persons in its member firms. Others may subscribe at the new rate of $5 per year or buy single copies at $1.25 each. Excepted are persons in companies eligible for regular ARF membership (U.S. advertisers, agencies and media), to whom the single-copy price remains $10.

solicits original papers. Reports of findings are favored over theoretical discussion. Manuscripts should be submitted in triplicate, double-spaced, with references, tables and figures on separate pages. Authors receive 50 free offprints. Letters of comment and criticism are also invited.

is intended for practitioners and users of advertising research. Limits on readers' time and journal space require that papers be as short as clarity permits.

is an open forum. Publication in it implies no endorsement of the writer's purpose, methods or views by the Advertising Research Foundation, its board of directors, or any of its committees.

is edited by the ARF Technical Staff:

Charles K. Ramond ..... Technical Director
Ingrid C. Kildegaard ..... Research Statistician
Gwyn Collins ..... Research Mathematician
Thornton C. Lockwood ..... Research Associate
Naomi Boretz ..... Librarian
Joey Harris ..... Secretary

Application to mail at Second Class Postage rates is pending at Brattleboro, Vermont

© 1961 by the Advertising Research Foundation, Inc.
3 East 54th Street, New York 22, N. Y.

Treatment of the Non-Response Problem[1]

TORE DALENIUS
University of Stockholm

"Survey estimates will never be unbiased—too many in the sample can't be reached." Dr. Dalenius deplores such defeatism and suggests three ways of handling non-response.

THE PURPOSE OF THIS PAPER is to present certain contributions to the treatment of the non-response problem in sample surveys. These contributions are described in the last three sections, against a background of the following general discussion.

Before it is possible to talk about a non-response *problem*, non-response must itself be defined. In this paper non-response will be discussed only in reference to the following situation.

A survey is to be made, using a plan whereby certain survey units, cited or otherwise specified (individuals in an opinion survey; households in a consumer survey), will be "observed" (for example, in such a way that an interview is made or certain data are collected by mail); but in reality only a proportion $r < 1$ of these units are observed.

According to this viewpoint, non-response may occur in a sample survey which is to be carried out by means of probability sampling, but not in a survey which is to be made by means of quota sampling, for example. Of course, knowledge of the distorting effect of non-response in probability sampling can be used in order to reveal, indirectly, the risk of bias in quota sampling.

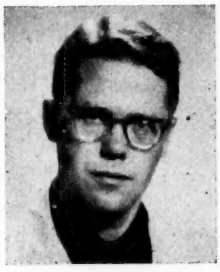
[1] This paper updates Chapter 11 of the writer's book, *Sampling in Sweden*, Stockholm: Almqvist & Wiksell, 1957; that chapter was based in turn on two prior papers (Dalenius, 1951 and 1955), also published only in Sweden.

For the following discussion, it is important to distinguish two causes for non-response:

1. Some units selected for the survey will not be duly contacted; a large portion of this category in interview surveys is made up of "not-at-homes."
2. Certain units will actually be contacted but either no observation is realized, for example because of refusal, inability to take part in the survey, etc., or an imperfect observation is realized that has to be rejected.

Effects of Non-Response

Nowadays it is generally accepted that a satisfactory survey design must contain a special plan for the treatment of the non-response problem. Otherwise it can seriously distort the survey results.



TORE DALENIUS is a docent at the University of Stockholm and a consultant to Tidningsstatistik (TS), an organization which does in Sweden what the ARF and the Audit Bureau of Circulations do in this country. From the University of Uppsala he received degrees in 1942 and 1948, and a Ph.D. in 1957. No stranger to the U.S., he was a research assistant at Cornell in 1947-48, studied sampling at the Bureau of the Census in 1951 and taught statistics at the University of California last fall as a visiting Fulbright professor. In Sweden he worked at the Gallup Institute from 1944 to 1947. Otherwise a civil servant from 1942 to 1956, Dr. Dalenius headed the Survey Research Center, a division of Sweden's Central Bureau of Statistics. He is a fellow of the American Statistical Association, the Institute of Mathematical Statistics and the American Association for the Advancement of Science. He has published over 70 papers on sampling theory and method.

Birnbaum and Sirken (1950) have noted the following. If in a survey with $100(1-r)$ per cent non-response, p per cent of units observed have a certain characteristic, the corresponding percentage which would have been obtained in a survey without non-response can be as great as p(max) or as small as p(min), where

$$p(max) - p(min) = 100(1-r)$$

The extensive literature of the non-response problem contains numerous empirical analyses. These have been made on the basis of "structure comparisons" (Crossley and Fink, 1951), comparisons of survey results from respondents observed on successive calls (Hilgard and Payne, 1944), direct analyses of non-response (Deming and Simmons, 1946) and comparisons of estimates based on the nights-at-home formula with estimates from call-backs (ARF, 1961).

These and other similar studies clearly show that there is ample reason to speak of a non-response *problem*, and that it must be dealt with in any particular survey.

Other things being equal, the smaller the non-response, the more exact the survey results will usually be. As Birnbaum and Sirken (1950) have shown, however, there is a lower limit under which it is not worthwhile to force the non-response; this limit can be said to give the optimal non-response. In practice one is usually far from reaching this limit unless one takes special pains. But participation enforced by means of compulsion *can* result in errors of reporting which have consequences more serious than those of non-response.

Measures Against Non-Response

Precautions which may be taken fall into two categories: *general* and *special*. The former apply to most situations, while the latter are intended for particular surveys.

General Measures. If respondents have a cooperative attitude toward surveys, this considerably simplifies the treatment of the non-response problem. In the U.S. the importance of making the public "survey-minded" was realized at an early date. Many methods have been used to create and maintain "survey-mindedness." Thus a special division of public relations, whose purpose was to prepare the way for census takers, was created for the U.S. census of 1940. With the aid of 2,000 local committees, an intense advertising campaign was conducted through various media. Similar measures were taken in the 1950 and 1960 censuses. All this has helped make the U.S. citizen relatively cooperative in furnishing information for the production of statistics.

Heneman and Paterson (1949) have studied the connection between the size of non-response due to refusals on the one hand, and the quality of the interviewers on the other. They found that with suitable training it was possible to decrease that part of the non-response caused by insufficient interviewer ability to about one-tenth of its size when untrained interviewers were used.

In training its representatives, the Survey Research Center of the Swedish Central Bureau of Statistics took into account results of this type. Lectures on the problem of non-response were given and both the central and field organizations were imbued with the idea that it could be overcome.

In 1953 the Swedish Survey Research Center included the non-response problem as a special item in its training program, having seen results of an analogous program. The Social Welfare Board had made a cost-of-living survey of households in built-up areas in 1949; non-response was about 35 per cent (Sveriges Officiella Statistik, 1953). In the opinion of the author, this high figure was largely due first to the fact that the non-response problem was then not considered serious in Swedish statistics production, and second to the sponsors' belief that it was impossible in this type of survey to reduce non-response to a very low level. As a result, the interviewers received insufficient training and indoctrination.

When the Social Welfare Board made a cost-of-living survey for rural districts in 1952, interviewers received much better training than in 1949, including a course where the non-response problem was specially treated. Non-response in this 1952 survey was held to eight per cent (Sveriges Officiella Statistik, 1955). Part of the decrease may be explained by a relatively lower resistance to surveys in rural districts than in towns. But the author sees in this decrease primarily a justification of an "it can be done" approach, and has since used this approach at the Survey Research Center and elsewhere.

Special Measures. The general measures above are seldom sufficient to solve the non-response problem in a particular survey; they must be complemented by special measures.

Swedish experience shows that the number of "absolute refusals" in an interview survey, for example, is relatively small: on the order of one per cent, according to the Swedish Survey Research Center studies in 1953-55. Non-response is caused

instead by difficulties in making contact with the respondent. One obvious method of reducing non-response is to make repeated callbacks. Deming (1953) has proved this to be highly worthwhile.

One difficulty in increasing the use of probability sampling in market surveys has been the difficulty in covering the extra costs, which usually accompany an attempt to achieve a small non-response, by a cost estimate given on a quotation basis. At the suggestion of the author, Tidningsstatistik AB applied the following "quotation technique" when requesting cost estimates from market survey organizations for two market surveys in 1953-54. The estimates were requested for the cost of a survey with the following conditions:

1. that it be based on probability sampling;
2. that its results be reliable with a standard error of 0.5 percentage points for an estimated proportion p = 15 per cent; and
3. that the amount paid to the organization undertaking the commission should be decided on a graduated scale; if the non-response were $100(1-r) = 15$ per cent, 100 per cent of the quotation amount would be paid; in the most favorable case the compensation would be 133 per cent of that amount, and in the most unfavorable case it would be 33 per cent.

In the first survey using the above plan, non-response fell to 11 per cent, a much better result than had normally been obtained in this field. In the second survey non-response was seven per cent.

For a survey done on a fixed budget it is frequently necessary to employ other special measures. Among them are Hendricks' correction method; the Politz-Simmons plan; and the Hansen-Hurwitz method.

Hendricks' correction method, discussed by Finkner *et al.* (1943), can be described as a method to reduce (or possibly eliminate) the *effect* of non-response.

For example, non-response in farm surveys often results in the over-representation of large farms. If there is a linear relation

$$X = A + BZ$$

between size Z of the farm and the quantity X to be studied in the survey, the unknown mean $\bar{X}$ is estimated by

$$\bar{x} = a + b\bar{Z}$$

where a and b are estimators of A and B respectively, and $\bar{Z}$ is the known mean of Z. The following is a method of estimating A and B on the basis of material distorted by non-response. The farms observed are divided into two groups, "small" and "large." For the respective groups, the means $\bar{z}_1$ and $\bar{z}_2$, $\bar{x}_1$ and $\bar{x}_2$ respectively are calculated, and B is now estimated by

$$b = \frac{\bar{x}_2 - \bar{x}_1}{\bar{z}_2 - \bar{z}_1}$$

and A by

$$a = \bar{x}_2 - b\bar{z}_2 \text{ or } a = \bar{x}_1 - b\bar{z}_1$$

The same method naturally lends itself to other types of populations. What is essential for the utilization of this method is a linear relation between X and Z. The method should, however, be regarded as a last resort.

An analogous idea is described by Hendricks (1949).

The basic idea in the Politz-Simmons plan, which will be described here as applied to an interview survey having persons as elements, was presented by Hartley (1946); Politz and Simmons (1949, 1950) developed the idea into a practical plan for use in interview surveys. In reference to priority see Deming (1953, footnote five).

The population is considered as being divided into strata which are homogeneous in regard to the probability P_{hj} of contacting the jth person during an attempted contact made at a time chosen at random (during that part of the day when contact is at all feasible). This probability is $P_{hj} = 1$ for persons who are at home all day; for persons who are away from home half of that part of the day when contact may be possible, it is $P_{hj} = \frac{1}{2}$, etc. A random sample of persons is contacted at a time chosen at random; 100 per cent of those for whom $P_{hj} = 1$ (on the average), 75 per cent of those for whom $P_{hj} = 0.75$, etc., and generally 100 P_{hj} per cent of those for whom the cited probability is P_{hj}, are contacted at that time. In the estimation of different parameters of the population, estimates of these P_{hj} values are used. If the completed questionnaires are transferred to punch cards, it is simplest to punch one card for respondents for whom $P_{hj} = 1$, two cards for respondents for whom $P_{hj} = \frac{1}{2}$, etc. From the theoretical point of view the Politz-Simmons plan is satisfactory; in practice it has been used by several marketing research firms.

Cornfield (1942) reported a method for solving the non-response problem in mail surveys, characterized by the fact that serious attempts to reduce the size of the non-response were concentrated on a random sample of non-respondents. Suppose for simplicity that a "simple random sample" comprising n units of some kind has been selected for a mail survey in order to estimate the popula-

tion mean $\overline{X}$. Without non-response, $\overline{X}$ would be estimated by the sample mean. However, the survey suffers from a non-response which comprises a proportion $1 - r$ of the whole of the initially selected sample. A random sample of non-respondents is selected, and thereafter this sample is interviewed personally. If $\bar{x}'_1$ is an estimator referring to the participants in the mail survey and $\bar{x}'_2$ refers to those units interviewed personally, the population mean is estimated by

$$\bar{x}' = r\bar{x}'_1 + (1 - r)\bar{x}'_2$$

In the article referred to above, Cornfield was content to give an unbiased method for the treatment of the non-response problem. The sample survey theory for this method was later given by Hansen and Hurwitz (1946). They showed that if one can indicate *a priori* approximately how large the non-response will be in the mail survey—this can often be done on the basis of previous experience—it is possible to fix the number of units to be contacted by mail and the number of non-respondents who will later be selected for a personal interview so that a given degree of precision is obtainable at minimum cost. A recent contribution of great practical interest is Frankel (1960).

Most attempts to treat the non-response problem lead to an extension of the time demanded by the survey; in many cases this is a serious disadvantage.

Three special measures designed by the author for dealing with the non-response problem are described in the sections which follow.

The Traveling Interviewer Problem

A prerequisite for using the Politz-Simmons plan is that the probabilities of finding the persons selected for the survey at home at a random call be determined for each person actually contacted. The data necessary for determining these probabilities are collected as part of the interview, i.e., they are estimated and made use of after the fact. In this section an approach will be outlined in which the probabilities of finding the persons at home at different times are made use of *before* the fact.

An interviewer is responsible for interviewing a number of persons spread over a certain "interview district." The assignment should be carried out as cheaply as possible within a fixed period of time. This calls for careful planning of the field work.

The difficulty of this planning task is largely due to the fact that different persons have different availability patterns; person A is likely to be at home during a period of time, e.g., between six and nine p.m., when person B is likely not to be at home.

The following practical procedure, to some extent used by the Survey Research Center of the Swedish Central Bureau of Statistics, is one efficient way of planning the field work. (This procedure may also increase efficiency in another way: by making the interviewer "efficiency-minded.") The field work is planned as follows:

1. The interviewer plots on a map the addresses of the persons making up his assignment.
2. Thereafter, he combines the addresses which are easily accessible to each other. Each such group corresponds to a sub-area.
3. Next, for each group separately, the interviewer analyzes the data relevant to the development of a rational plan for the field work. Such data are used for the composition of work-loads in terms of variables that forecast whether a person is at home or not at different times, e.g., sex, age, economic activity, industry (agricultural—non-agricultural), residence (urban-rural). In Sweden these data are available in the population registers used as frames and may therefore be given to the interviewer.
4. If a work-load is small enough to require only one successful visit to the sub-area, the interviewer picks the time likely to be best, i.e., corresponding to a large chance of contacting all persons of this work-load.

   If a work-load is large, the interviewer splits it into parts as homogenous as possible with respect to variables determining the probability of finding a person at home at different times. These parts should be of manageable sizes.

   Finally, the solutions arrived at for each such work-load or part of work-load should be analyzed simultaneously for possible economies from combined scheduling of the visits.

The combination of addresses into groups as discussed above obviously means the use of "sub-optimization."

The data necessary to apply this kind of strategy may be collected in connection with interview surveys, both as statistics and as experience on the part of the interviewer. Such data should prove especially helpful in panel surveys.

Conceptually, one may associate with each person a set of probabilities for his being at home at different points of time; this set defines the person's "availability pattern." The difficulty the interviewer faces in the field is largely caused by two factors: the availability patterns are different for different persons, and the probabilities are zero for many points of time.

It may be pointed out that if all probabilities are equal to one, the problem becomes the classic traveling salesman problem.

To be realistic, a theoretical set-up should prove to be rather complicated. Among other things, the probabilities $P_j(t_i)$ for the jth person being at home at point of time t_i, where $i = 1, 2$, etc., refer to events that are not independent. A Markov chain which is characterized by the fact the $P_j(t_i)$ depends only on $P_j(t_{i-1})$ may prove to be unrealistic. Models that

involve a more realistic dependence are likely to be too complicated to be of practical use.

The difficulty of the problem is enlarged by the fact that the field work has to be finished at a certain deadline; this may make it necessary to accept a small risk of getting $(1-r)n$ non-responses.

Subsampling A Priori

In most mail surveys a good deal of time passes before the response rate is high enough to obtain the result.

If a speedy preliminary estimate is needed, the following may be considered. As the sample for the mail survey is being selected, a subsample of size n′ is selected from it. All available resources are immediately concentrated on obtaining a relatively high response rate from this subsample, after which the preliminary estimate is made.

The simplest way administratively of completing the survey is to base the final estimate on the $n-n'$ units remaining; this means that the n′ units are rejected, which can be justified if n′ is relatively small or if the cost of collecting data is negligible. To use both the n′ and the $n-n'$ units, it is necessary—in order to avoid bias—to eliminate all spontaneously participating units in both sample parts counted from the date when the n′ units were made the object for special attention to increase the response rate (Hansen, 1956).

Subsampling in the Field

Here the method described is aimed at that part of the non-response problem which depends on the fact that certain units selected for a survey are hard to contact. The method is especially useful in interview surveys and has been successfully applied by a Swedish marketing research firm. The method will be described here as applied to an opinion survey.

It is characteristic of this application that the field work is administered in such a way that two different sampling fractions f_P and f_A (where $f_P > f_A$) are used for the (ultimate) selection of respondents who are found at home (P = present) at the first call and absent (A = absent), respectively.

Often an opinion survey is designed in the following way:

1. a sample of households is selected (in a nationwide survey within a set of primary sampling units);
2. the interviewer makes a complete list of potential respondents for each household selected;
3. from this list of potential respondents the interviewer selects *one* for interview.

The procedure worked out by Kish (1949) is suitable for the selection of a respondent. According to this procedure the potential respondents in the jth household are allotted selection numbers 1, 2, . . . Q_j according to some special rule, whereby one of these potential respondents is selected with the help of a special table for random sampling, so constructed that every potential respondent has the chance $1/Q_j$ of being selected. The following presentation is based in principle on Kish's method.

The proposed method means:

1. that a sample of households is selected;
2. that the interviewer makes a complete list of potential respondents for each household selected; and
3. that the Q_j potential respondents in the jth household are given selection numbers which are chosen in such a way that the potential respondents who are present and in consequence can be interviewed immediately if selected, get a greater chance of being selected than those who are absent.

Here the method—which may be characterized as a special design of the Hansen-Hurwitz scheme—will be illustrated for the special case when $f_P = 2f_A$, i.e., those present are given twice as large a chance of being selected as those absent. The interviewer begins by making a list of all potential respondents belonging to the household, beginning with the head. He notes the day of the month when the person in question was born and indicates by P or A whether the person in question is present or absent. A result of these operations would be the following list:

Potential Respondent	*Position in Household*	*Day of Birth*	*P, A*
Andersson, Axel	Head	20	P
Emma	Wife	12	P
Erik	Son	15	A
Elsa	Daughter	3	P

These four persons are now allotted selection numbers in a randomized fashion. The numbering begins say with the persons having the lowest day of birth, etc.; other practically applicable devices for randomization may be preferable. Those coded "P" receive *two* consecutive numbers, those coded "A" *one* number in the series 1, 2, . . . Q_j. . . .

In this instance the result is

Potential Respondent	*Position in Household*	*Day of Birth*	*P, A*	*Selection Number*
Andersson, Axel	Head	20	P	6, 7
Emma	Wife	12	P	3, 4
Erik	Son	15	A	5
Elsa	Daughter	3	P	1, 2

Now a person is selected from the household with a mechanism that gives every P-person $\frac{2}{7}$ and every A-person $\frac{1}{7}$ chances of being selected. Such a mechanism is Kish's selection table constructed for a

household consisting of seven potential respondents. It is apparent that the number interviewed at the first contact in this way will be significantly larger than otherwise.

The selection of one and only one respondent from each household independent of the size of the household can be motivated by one or other (possibly both) of the following reasons: first, the intraclass correlation between persons within the household is high and the cost of additional interviews is also high, so that it would not pay from the standpoint of "efficiency." Second, an interview with one person would influence interviews with other persons in the same household.

An advantage of not specifying that exactly *one* person is to be selected from each household is that the tabulations can be made simpler. When exactly *one* person is selected from each household the interview must be "weighted" by the reciprocal of the probability with which the person concerned has been selected, which may mean using between five and ten weights.

If the condition that the sample must include exactly *one* person from each household is withdrawn, the special technique referred to above can be applied in one of the two following ways which are here referred to as variant I and variant II, respectively.

Variant I. The interviewer makes up a running list of the potential respondents and assigns them selection numbers in the same manner as above. Those selection numbers which determine the actual respondents are stated in advance on the list. Such an application would give the following result:

Household Number	*Potential Respondent*	*Position in Household*	*Date of Birth*	*P, A*	*Selection Number*
1	Bergstrom, Karl	Head	8	A	3
	Anna	Wife	27	P	4, 5
	Elsa	Daughter	5	P	1, 2
2	Andersson, Axel	Head	20	P	11, 12
	Emma	Wife	12	P	8, 9
	Erik	Son	15	A	10
	Elsa	Daughter	3	P	6, 7

Variant II. The interviewer makes up two lists, a P-list which includes the potential respondents present and an A-list which includes those absent. Each such list is made analogous to that used in variant I. On the P-list the potential respondents are numbered (where $f_P = 2f_A$ which is used here as an illustration) with "double numbers:" 1-2, 3-4, 5-6, etc., and on the A-list with "single numbers:" 1, 2, 3, etc. Then from each list "every nth" is applied to the number series according to the description given for variant I.

Clearly, these techniques assume that the interviewers work in an honest way; this aspect is, however, by no means specific to these techniques.

In the example above the use of the special technique has been illustrated with $f_P = 2f_A$. However nothing prevents the use of any relation $f_P = tf_A$, where t need not be an integer. For example, if t = 1.5, the numbering system can be of the type 1-2-3 for P-persons and 1-2 for A-persons.

The procedure outlined above represents "double sampling with stratification" according to the terminology of Hansen *et al.* (1953). In this reference, the basic theory is given for determining the sampling fractions f_P and f_A in an optimum fashion.

The method described above is of course not limited to the type of applications which have been used here as illustrations. The method can, administratively speaking, be applied to the second step of the field work in the opinion survey illustrated, that is, in the interviews with the A-persons initially selected for interview. Further, the method can be applied to the selection of households in a survey. Consider a survey using area sampling, the design of which calls for a sample of geographically compact segments comprising about ten households. The interviewer may list the households in each segment and then interview a sample of these households. He may then be instructed to work according to variant II given above, for example.

Finally, the possibility of combining this method with other procedures which also aim at increasing the information obtained per unit cost should be mentioned. One example will be given, Hansen (1955):

> A person at home, but not in the sample, could be asked the questions normally required of the survey respondent who is not at home, but is in the sample. If the questions are of a type which would make for a high correlation between this second-hand answer and the survey respondent's answer, it should be possible to stratify the not-at-homes on this second-hand answer and to use sampling fractions for the resulting strata which would result in smaller sample sizes for the not-at-homes. The estimation procedure, for the not-at-home component, would take advantage of the second-hand answer in a ratio or in a regression type of estimate.

REFERENCES

ADVERTISING RESEARCH FOUNDATION. *A Comparison of Estimates from the Nights-at-Home Formula with Estimates from Six Calls.* New York: Advertising Research Foundation, Inc., 1961.

BIRNBAUM, Z. W. AND M. G. SIRKEN. On the total error due to non-interview and to random sampling. *International Journal of Opinion and Attitude Research,* 1950, pp. 179-191.

CORNFIELD, J. On Certain Biases in Sampling of Human Populations. *Journal of the American Statistical Association*, 1942, pp. 63-68.

CROSSLEY, H. M. AND R. FINK. Response and Non-Response in a Probability Sample. *International Journal of Opinion and Attitude Research*, 1951, pp. 1-19.

DALENIUS, T. The Problem of Non-Response in Statistical Surveys. In *The Swedish Market*, 1951.

DALENIUS, T. The Problem of Not-at-Homes. *Statistisk Tidskrift*, 1955, pp. 208-211.

DEMING, W. E. On a Probability Mechanism to Attain an Economic Balance between the Resultant Error of Response and the Bias of Non-Response. *Journal of the American Statistical Association*, 1953, pp. 743-772.

DEMING, W. E. AND W. SIMMONS. On the Design of a Sample for Dealers' Inventories. *Journal of the American Statistical Association*, 1946, pp. 16-33.

FINKNER, A. L., J. J. MORGAN AND R. J. MONROE. *Methods of Estimating Farm Employment from Sample Data in North Carolina, Bulletin No. 75*. Raleigh: North Carolina Agricultural Experimental Station, 1943.

FRANKEL, LESTER R. How Incentives and Subsamples Affect the Precision of Mail Surveys. *Journal of Advertising Research*, Vol. 1, No. 1, September 1960.

HANSEN, M. H. Personal correspondence, 1955.

HANSEN, M. H. Personal correspondence, 1956.

HANSEN, M. H. AND W. N. HURWITZ. The Problem of Non-Response in Sample Surveys. *Journal of the American Statistical Association*, 1946, pp. 517-529.

HANSEN, M. H., W. N. HURWITZ AND W. G. MADOW. *Sample Survey Methods and Theory* (Vol. 1). New York: John Wiley and Son, 1953.

HARTLEY, H. O. Discussion quoted in: Yates, F. A Review of Recent Statistical Developments in Sampling and Sampling Surveys. *Journal of the Royal Statistical Society*, 1946, pp. 12-43.

HENDRICKS, W. A. Adjustment for Bias by Non-Response in Mailed Surveys. *Agricultural Economic Research*, 1949, pp. 52-56.

HENEMAN, G. H. AND D. G. PATERSON. Refusal Rates and Interviewer Quality. *International Journal of Opinion and Attitude Research*, 1949, pp. 392-398.

HILGARD, E. R. AND S. L. PAYNE. Those Not at Home: Riddle for Pollsters. *Public Opinion Quarterly*, 1944, pp. 254-261.

KISH, L. A Procedure for Objective Respondent Selection within the Household. *Journal of the American Statistical Association*, 1949, pp. 380-387.

POLITZ, A. AND W. SIMMONS. An Attempt to get the "Not at Homes" into the Sample without Callbacks. *Journal of the American Statistical Association*, 1949, pp. 9-31.

POLITZ, A. AND W. SIMMONS. An Attempt to get the "Not at Homes" into the Sample without Callbacks. *Journal of the American Statistical Association*, 1950, pp. 136-137.

SVERIGES OFFICIELLA STATISTIK. Levnadskostnaderna i tatortshushall ar 1948. Stockholm.

SVERIGES OFFICIELLA STATISTIK. Levnadskostnaderna pa landsbygden ar 1951. Stockholm.

ADDENDUM

In my review article, "Sales and Advertising of Lettuce," which appeared in the March 1961 issue of this *Journal*, I used a case study as an illustration of some of the major concepts involved in quantitative measurement of the impact of advertising. I cited a study conducted by "T. A. Mainwaring *et al.*," of Stanford Research Institute as the source of certain data. The *"et al."* reference does not adequately acknowledge the work of Professor H. S. Houthakker, Harvard University (formerly Stanford University), who served as a consultant to Stanford Research Institute on the lettuce study. He conducted the regression analysis the results of which are shown in Tables 1 and 2 of my article. A large portion of my review and many of the suggestions for future work on advertising evaluation were, of course, based on and inspired by Professor Houthakker's original findings.

—FRANK MEISSNER

ARF 7th Annual Conference Program
Hotel Commodore, New York City, October 3, 1961

9:30 A.M. MAKING ADVERTISING DECISIONS

Charles E. Swanson, Assistant Director of Research, The Curtis Publishing Company; Chairman, ARF Seventh Annual Conference Program Committee

9:45 A.M. NEW TECHNIQUES IN VISUAL RESEARCH

Presiding: Leonard Kent, Vice President and Director of Research, Needham, Louis and Brorby, Inc.; Vice Chairman, ARF Seventh Annual Conference Program Committee

Can the Visual Effectiveness of Advertisements Be Controlled?

Robert Stover, Manager, Human Factors Branch, HRB-Singer, Inc.

Discussant: Alfred A. Whittaker, Vice President and Director of Advertising, Bristol-Myers

Intrigue in Advertising: Motivational Effects of Visual Organization

Clark Leavitt, Creative Research Supervisor, Leo Burnett Company, Inc.

Discussant: David B. McCall, Senior Vice President and Copy Chief, Ogilvy, Benson & Mather, Inc.

11:15 A.M. REPORT ON AN ADVERTISING RESEARCH TECHNIQUE

Commercial Research Organization Submitting Winning Paper in ARF Competition

12:30 P.M. LUNCHEON

Presiding: Arthur Hull Hayes, President, CBS Radio, A Division of Columbia Broadcasting System, Inc.; Chairman of ARF Board of Directors

Keynote Address

The Honorable Edward Gudeman, Under Secretary of Commerce

2:30 P.M. CONCURRENT SPECIAL INTEREST GROUP SESSIONS

EVALUATING SPECIAL MARKET MEDIA

Presiding: John C. Spurr, Marketing Research Consultant, McGraw-Hill Publishing Company, Inc.

Better Research—A Job for Both Advertiser and Publisher, Harry Waddell, Executive Vice President, McGraw-Hill Publishing Company, Inc.

Discussant: John Veckly, Director of Advertising, United States Steel Corporation

New Findings from Scott's Special Advertising Research Study, Burt B. Roens, Industrial Marketing Manager, Scott Paper Company

Discussant: James C. Becknell, Jr., Research Psychologist, E. I. du Pont de Nemours & Company, Inc.

Evaluating Industrial Advertising—Realistically and Objectively, Richard C. Christian, President, Marsteller, Rickard, Gebhardt and Reed, Inc.

Discussant: George M. Robertson, Manager, Advertising Services and Measurement, Advertising and Sales Promotion, General Electric Company

EVALUATING MASS MEDIA

Presiding: John D. Henry, Manager, Market Research Department, The Procter & Gamble Company

Pitfalls and Potentials in Evaluating Mass Media, Harry V. Roberts, Professor of Statistics, School of Business, University of Chicago

Discussant: Alfred N. Watson, Vice President, Marketing Services, United States Rubber Company

Measuring the Relative Sales Effectiveness of Media, George H. Brown, Director, Marketing Research Office, Ford Motor Company

Discussant: William M. Weilbacher, Vice President, Dancer-Fitzgerald-Sample, Inc.

Methods of Evaluating Television Advertising Effectiveness, Jack B. Landis, Director of Developmental Research, J. Walter Thompson Company

Discussant: Lee Rich, Senior Vice President in Charge of Media and Programming, Benton & Bowles, Inc.

A Comparison of Eight Audience Estimates[1]

J. STEVENS STOCK
J. Stevens Stock Research Company

Magazine audiences are usually estimated from a sample's answers to questions. The writer finds few differences among estimates obtained by eight different questions.

PRINT MEDIA MEASUREMENTS start with ABC figures, the numbers of copies sold. Though quite accurate, they have nothing to do with quality. For this reason, the industry has tried to develop audience measurements which qualify the ABC numbers. Some of these are: the number and character of people who read an average issue of the magazine; the frequency with which various readers peruse the magazine over several issues; and the behavioral characteristics of readers' reading of the magazine—time spent reading, times picked up in a week, pages opened, and so forth. These are difficult to measure because they are not simple, observable, countable things, like copies of the magazine. They are measures of what people do, and this we cannot readily observe. All we can usually observe is what people *say* they do.

Fortunately, there is a connection between what people in large numbers say they do and what they actually do. This connection is perhaps closer than recent writers on the subject have conjured from their studies of abnormal behavior. At any rate, we are beginning to accumulate a body of knowledge about how a group's average behavior relates to the average of its verbal responses to interviewers' questions.

There are three classes of reasons why a person's reported behavior may differ from his actual behavior:

1. *Forgetting.* This is the fading process, analogous to entropy in the physical sciences. Our knowledge cools off or decays over time.
2. *Perception.* Things important to us look bigger to us, regardless of how we define importance. Viewed in time, important things seem more recent; unimportant things are recalled as happening a longer time ago. The tendency is to recall pleasant, more respectable things, as happening more recently and frequently than the unpleasant, less respectable things.
3. *Motivation.* All men have desires—some secret, some subconscious. When we try to recall what we did, we tend to recall what we think we *should* have done or in some sense would *like* to have done. The effect is to respond to questions in a way that characterizes us most favorably.

J. STEVENS STOCK is president of his own research firm. He received his advanced education in mathematics and statistics at the American University in Washington, D. C., where he taught in the graduate school. He has also taught in the graduate schools of Princeton University, the University of Chicago and the City College of New York. During the war he served as special assistant to the Secretary of the Navy with the rank of senior commander, and later as director of research at McCann-Erickson in Chicago and vice president of Alfred Politz Research, Inc. He is the author of books and articles on the mathematics and psychology of research in marketing and public relations.

[1] *Newsweek* magazine underwrote this study in order to evaluate, in a systematic way, the methodological considerations described.

It is commonly recognized that we should take these things into account when we go about asking people to tell us about their reading. Insofar as we can, our results will yield answers that come closer to describing the actual behavior of the people we study.

Out of these problems we can derive three obvious rules. Ask for specific answers, free of judgment on the part of the respondent. Don't tax his memory. Don't require answers he might think would characterize him. Much that is poor in media research can be attributed to disregard of one or more of these rules.

Because there had been so many poor methods of trying to infer people's actual reading behavior from their oral answers to memory questions, some years ago Louise McCarty DuBois, working for *Life,* developed a new method. In the DuBois method the respondent was first taken completely through a particular issue of the magazine and then asked if he was sure he had previously read it.

Later, *Life* financed studies by Alfred Politz Research of variations of this method. The result was a technique we may call the Recognition Method. It has proven to be precise, and has been adopted generally as the standard method of measuring readership. The success of this standard method is largely ascribable to its adherence to the three rules of question construction mentioned previously.

In the Recognition Method as described by Politz, "a reader of an issue is defined as a person exposed to any of the issue's editorial contents. The sum of readers by this definition is called the 'Audience.' Operationally, a person is counted as a 'reader' only if he says he is certain he has looked into the issue before, after having examined the issue's editorial contents." (Politz, 1957, p. 184.)

In effect, this method measures "the audience of an average issue"—the average of the issues used during the interviewing period. But the Recognition Method involves certain problems:

1. It requires personal interviewing; it cannot be done by mail or telephone.
2. The interviewer must carry copies of the magazines whose audiences are being measured. These copies must be four to five weeks old. If the interviewing takes several weeks, the magazine must be replaced.
3. The magazines must be altered to eliminate much of the advertising material and show mainly the major editorial parts.
4. The interview is so long that relatively few magazines can be covered in any one interview.
5. Interviewers carrying magazines are often suspected of using the survey quiz as a ruse to sell subscriptions. This obviously leads to refusals.

As a result, the Recognition Method is so expensive that only large circulation magazines can use it, and they only infrequently.

Recognition vs. Recall

This paper reports an experimental study of a simpler, more practical method than the Recognition Method. It compares the number of readers found, using it and several versions of what we here term the Recall Method.

Most recall questions are aimed at estimating the number of readers in some average period of time, as opposed to the Recognition Method of estimating the number of readers of an average issue. It may not be obvious that the number of readers of an average issue of a weekly magazine is very close to the number of people who read any issue of it in an average week.

For instance, some people read the same issue in two separate weeks. They would be counted twice in weekly readership but only once in issue readership.

On the other hand, some people read two different issues in the same week. They would be counted twice in issue readership but only once in weekly readership.

Continued reflection along these lines suggests, however, that although differences might exist between the two methods for specific issues, or specific weeks for a weekly magazine, there tend to be no differences between measures of *average* issues or *average* weeks.

For example, let us compare the number of times any issue of a weekly magazine is picked up and read in an average week, with the number of times an average issue is picked up and read. If we think of last week as an average week and last week's issue as an average issue, then the number of pick-ups last week represents the pick-ups of all previous issues still available to be picked up. Thus an average week's pick-ups is a count of all past issues. Similarly, last week's issue will stay around for many future weeks to be picked up in future weekly pick-ups. Therefore, the number of pick-ups in an average week is identical with the number of pick-ups of an average issue.

The number picking up an issue of a weekly at least once defines the issue audience; the number picking up any issue at least once during the past week is the weekly audience, and these two audiences—issue and weekly—are equal. Specifically, the proposition is as follows.

since

Pick-ups of *all* issues during *a given week*
equals
Pick-ups of *a given* issue during *all* weeks.

It follows that

Persons reading *all issues at least once in a given week*
equals
Persons reading *a given issue at least once* in *all* weeks.

In this study we seek to learn if there is any measurable difference between the average weekly audience as estimated by various recall methods, and the average issue audience as estimated by the Recognition Method.

Practically all methods we know of, other than the Recognition Method, are recall methods in the sense that they rely upon memory unaided by the presence of the issue, of recent specific reading or reading habits.

Recall questions on reading *habits* are extremely difficult to accept on the theoretical grounds discussed earlier. Take the following question, still used in some audience research:

About how often would you say you actually read *(magazine)*?

This question, if it yielded accurate answers, would produce good estimates of reading in an average period of time. The average weekly audience would include all people who say "every issue" or "every week" plus *half* the people who say they read every *two* weeks, plus a *third* of the people who say they read every *three* weeks, and so on.

But "recall-of-habit" questions like this do not easily yield clear-cut answers that can be manipulated in the above formula. The main problems are:

1. Answers are not specific. How do we count people who say "frequently," "sometimes," "once in a while?"
2. The question is a puzzler to the conscientious respondent. It entails this sort of a memory feat: "How many times have I read that magazine in the past year? Twenty times. Twenty divided into fifty-two is 2.6. I read it every 2.6 weeks"!
3. Some respondents will unconsciously answer as if the question might tend to characterize them. We can safely assume that many people (we will perhaps never know how many) feel that regular reading of some magazines carries more prestige than regular reading of others.

Recent specific reading is easier to recall than general reading habits. Here is the question most commonly used to measure it:

When was the last time (if ever) you read (or looked into) a copy of _____?

This question has relatively few difficulties, either theoretical or practical. It conforms much more closely to our three rules of good questioning, to wit:

1. The question is specific. The only exception to this is the relatively few people who answer "last week." Such respondents have to be questioned about the day of their reading experience. For instance, on a Saturday interview, the respondent may refer to a calendar week and really mean that he last read a magazine as far back as a week ago last Sunday—fourteen days ago, counting today as one day ago.
2. The question does not tax the memory; only one specific event, the last one, is recalled. If that last event was many months ago, an error of a month or so either way does not materially affect our estimate of the number of readers in an average week or two weeks.
3. The question does less characterizing of the individual. What one happened to do or not to do on one recent occasion does not necessarily reveal continuing good or bad habits.

METHOD

Sample

Ninety blocks in an upper income suburb of Philadelphia were selected. The reason for choosing an upper income area was to locate respondents with a potentially high incidence of magazine reading. The resulting high percentages, being relatively more stable in small samples, are considerably more precise for comparisons when trying to measure differences associated with various questioning and interviewing methods.

Starting on the southwest corner of each of the 90 blocks, the first three households were selected. Adjacent houses were chosen in order to simplify reaching them, to facilitate respondent identity in call-backs, and so that the three samples of households (one from each block) would in a sense be comparable but independent. Within each household, from all persons over 12 years of age, one was chosen at random (by the standard X-ing pattern method) for interview and reinterview.

Interviewing

Personal interviews were conducted with two of the three respondents in each block; the third was interviewed by telephone.

One personal interview questionnaire followed in detail the standard Recognition Method. The magazines used were never less than four or more than five weeks old.

The other personal interview questionnaire asked the recall question quoted above, as well as certain scaling questions used by Sindlinger and Company. It went further, however, to probe on succeeding questions persons systematically dropped out by Sindlinger's scale (Stock, 1959).

The telephone interview questionnaire was as nearly identical to the personal recall questionnaire as possible, given the intrinsic characteristics of the two types.

During the week of October 31, 1960, Wave I interviews were conducted with each of the three randomly selected respondents, using one or the other of the three questionnaires. The order of the questionnaires on each block was randomized among the persons selected in the first, second and third households.

Six weeks later, during the week of December 12, 1960, the *same* respondents were reinterviewed in Wave II. One-third of those interviewed on Questionnaire A on Wave I were reinterviewed on Questionnaire A; another third on Questionnaire B; and the final third on Questionnaire C. Similarly, those interviewed on Questionnaires B or C in Wave I were divided into thirds to be interviewed on Questionnaires A, B and C in Wave II. Thus there were nine sequences of questionnaires in the first and second waves of interviewing: AA, AB, AC, BA, BB, BC, CA, CB and CC.

Eight weeks after Wave II, in the two week period beginning February 6, 1961, each person was *again* reinterviewed in Wave III. Each of the nine groups was further sub-divided into three groups, making a total of 27 sequences of questionnaires in the three waves. Table 1 shows the completion rates by waves.

TABLE 1

NUMBER AND PER CENT OF COMPLETED INTERVIEWS

	Wave I (October)		*Wave II (December)*		*Wave III (February)*	
Method	*No.*	*Per Cent*	*No.*	*Per Cent*	*No.*	*Per Cent*
Personal Recognition	90	100	79	88	69	77
Personal Recall	89	99	74	82	74	82
Telephone Recall	88	98	86	96	70	78

In addition, as a fourth wave, all 270 persons in the sample were sent a mail recall questionnaire March 27, 1961. Returns after one follow-up were 244 or 90 per cent.

Six weekly magazines were studied in each questionnaire: *Life, Newsweek, Saturday Evening Post, Sports Illustrated, Time* and *U. S. News & World Report.*

All interviewers were personally trained and the same interviewers were used in all three waves. To minimize possible conditioning, each of the three interviews with a respondent was conducted by a different interviewer than had contacted him before. To overcome refusals, especially in Wave III, a few cases were allowed where an interviewer who had been successful on an earlier wave was used to call back a second, third, or fourth time in an effort to gain cooperation.

RESULTS

The results are summarized in Table 2. Recall of reading during the past week, even when measured by telephone, gave very nearly the same numerical size of audience as the standard Recognition Method.

None of the methods showed appreciable interaction with the magazines studied. Where one method overstated a particular magazine relative to another method, it overstated all magazines in about the same proportion.

The mail questionnaire (recall of reading within the past *two* weeks only), as applied here with follow-up and high return, gave roughly the same answers as the two week Telephone Recall Method and the two week Personal Recall Method. Even the first returns before follow-up were not significantly different from the telephone and personal interview Recall Methods.

Specific comparisons follow.

TABLE 2

PER CENT CLAIMED READING, ALL WAVES COMBINED

	Methods							
		Personal Recall		*Telephone Recall*		*Mail Recall*	*Scaling Question*	
Magazine[1]	*Personal Recognition*	*One Week*	*Two Week*	*One Week*	*Two Week*	*Two Week*	*First Wave Only*	*"How Often Read" Question*
Life	59.6%	61.5%	70.0%	55.3%	59.4%	70.9%	56.2%	60.0%
Newsweek	25.2	26.2	30.0	24.6	28.7	25.4	27.5	27.6
Saturday Evening Post	45.0	43.0	53.2	41.0	48.4	52.0	47.0	43.6
Sports Illustrated	8.0	13.9	16.5	13.5	16.0	16.4	13.8	14.2
Time	36.6	32.8	39.7	33.6	36.5	44.7	28.7	35.9
U.S. News & World Report	14.3	19.8	21.1	14.3	16.4	18.0	12.6	18.4
(Base)	(238)	(237)	(237)	(244)	(244)	(244)	(87)	(181)

[1] It seems hardly necessary to point out that these estimates, coming as they do from a small sample of one suburb, are not intended to represent the national audiences of these magazines, and cannot possibly reflect any such audience except by chance.

Conditioning

Conditioning showed no significant effect. Table 3 shows the combined telephone plus personal interview one week recall estimates for each of the three waves.

TABLE 3

PER CENT CLAIMED READING, ONE WEEK RECALL, PERSONAL AND TELEPHONE COMBINED

	Wave		
Magazine	*I*	*II*	*III*
Life	59.3%	58.1%	57.6%
Newsweek	27.7	23.8	24.3
Saturday Evening Post	39.5	42.5	44.4
Sports Illustrated	12.4	11.8	17.4
Time	29.9	34.4	36.1
U.S. News & World Report	16.4	19.4	15.3
(Base)	(177)	(160)	(144)

Indeed, neither method of interviewing nor conditioning significantly affected recall estimates of magazine audiences. The complete analysis of variance of the six estimates of each magazine made by the two interviewing methods in the three waves is shown in Table 4.

TABLE 4

ANALYSIS OF VARIANCE OF ONE WEEK RECALL ESTIMATES

Source of Variation	*Sum of Squares*	*Degrees of Freedom*	*Mean Square*	*F*
Magazines	8285	5	1657.0	56.2*
Methods (Personal vs. Telephone)	38	1	38.0	1.3
Waves	12	2	6.0	<1
Interactions:				
Magazines—Methods	74	5	14.8	<1
Magazines—Waves	114	10	11.4	<1
Methods—Waves	35	2	17.5	<1
Magazines—Methods—Waves	295	10	29.5	
Total	8853	35	——	

* Significant beyond .05 level.

Telephone vs. Personal

There was no significant difference between estimates by telephone and personal interviewing to measure one week recall, where all conditions of sampling rigor were equal. The results for the three waves combined are shown in the second and fourth columns of Table 2.

Recognition vs. Recall

Table 5 shows a total analysis of variance of the nine separate estimates of the average audience of each magazine made in three waves by the three methods: Personal Recognition, Personal Recall and Telephone Recall. From it we conclude that whether one measures recognition or recall, by personal or telephone interviewing, the results are essentially the same, even under conditions of reinterviewing.

TABLE 5

ANALYSIS OF VARIANCE OF ONE WEEK RECALL ESTIMATES

Source of Variation	*Sum of Squares*	*Degrees of Freedom*	*Mean Square*	*F*
Magazines	13,855	5	2771.0	161.9*
Methods (Personal and Telephone combined *vs.* Recognition)	37	2	18.5	<1
Waves	68	2	34.0	1.25
Interactions:				
Magazines—Methods	216	10	21.6	<1
Magazines—Waves	135	10	13.5	<1
Methods—Waves	83	4	20.8	<1
Magazines—Methods—Waves	543	20	27.2	
Total	14,947	53	——	

* Significant beyond .05 level.

One Week vs. Two Week

Table 6 shows that two week recall estimates of audience overstated one week recall estimates by about 15 per cent. This seems obvious: all the people who have read a magazine in the past week are included in those who have read it in the past two weeks.

TABLE 6

PER CENT CLAIMED READING, PERSONAL AND TELEPHONE RECALL COMBINED, THREE WAVES COMBINED

	Recall Estimates		
Magazine	*One Week*	*Two Weeks*	*Per Cent Two Exceeds One*
Life	58.2%	64.5%	10.8%
Newsweek	25.4	29.3	15.3
Saturday Evening Post	42.0	50.7	20.7
Sports Illustrated	13.7	16.2	18.2
Time	33.3	38.0	14.1
U.S. News & World Report	17.0	18.7	10.0
(Base)	(481)	(481)	

The overage of the two week recall method is, obviously, all people who have not read in the past seven days but who did read in the period eight to 14 days ago. The column at the right in Table 6 shows that while the average overstatement is 15 per cent, there is some variation around this average, from an average 10 per cent for *U. S. News & World Report* to an average 21 per cent for *Saturday Evening Post*. This interaction between time of recall and magazine was statistically significant but not appreciable.

It can be shown that the two week recall method really is an estimate of the *two week net or cumulative audience*—a statistic exactly comparable to the net or cumulated issue audience. Therefore, magazines which cumulate audiences faster than others will tend to have a slightly higher two week recall estimate than magazines that cumulate audiences at a slower rate.

The Recall Method estimates of net or cumulative audiences are shown in Table 7. As we thought in our theoretical discussion, these data tend to be erratic in long time recall; it is hard to place an event accurately that occurred several months ago. But these data can easily be smoothed by plotting the cumulated frequencies against the logarithm of time since last reading. The result is very nearly a straight line.

TABLE 7

PER CENT CLAIMED READING, PERSONAL AND TELEPHONE RECALL COMBINED, ALL WAVES COMBINED (BASE = 144)

Recall Span: "Read Last . . ."	*Life* %	*News-week* %	*Sat. Eve. Post* %	*Sports Illus.* %	*Time* %	*U.S. News & World Report* %
Up to 7 days ago	58.2	25.4	42.1	13.7	33.3	17.0
8 to 14 days ago	6.3	3.9	8.7	2.5	4.8	1.7
15 to 21 days ago	5.4	3.7	2.9	2.1	4.4	3.7
22 to 30 days ago	4.8	8.1	5.6	6.2	7.5	4.2
31 to 60 days ago	4.4	4.8	3.3	2.3	2.9	4.0
61 to 90 days ago	2.5	1.7	2.3	1.5	2.5	2.5
91 to 182 days ago	3.5	7.3	5.8	4.8	6.4	7.7
183 to 273 days ago	0.4	0.6	0.8	0.4	0.2	0.2
274 to 365 days ago	0.0	0.0	0.2	0.2	0.4	0.0
"About a year ago"	2.4	10.0	6.5	6.0	5.8	5.4
"Over a year ago"	3.7	7.1	5.4	6.9	8.3	8.1
"Never read"	8.4	27.4	16.4	53.4	23.5	45.5

The Scaling Method

This method of determining readership has been described previously (Stock, 1959). As shown in Table 2, estimates from it tend to agree closely on the average with the Recognition and one week Recall data. Table 8 presents the estimates from the Scaling Method (first wave, telephone interviewing only) compared with the Recognition Method, all waves, and the one week Recall Method, all waves, personal and telephone interviewing combined.

In two important ways, these Scaling Method estimates are not derived in exactly the same way as done by Sindlinger and Company. These estimates are based on a very rigorous random selection of individuals within the household, while the Sindlinger practice is somewhat more haphazard, with the result that people more apt to answer the telephone are over-represented. This tendency, of itself, has been shown to increase the likelihood of finding magazine readers. Moreover, we had to exercise considerable probing to make some of the Sindlinger questions yield precisely codable answers.

TABLE 8

PER CENT CLAIMED READING, BY THREE METHODS

Magazine	*Recognition, All Waves Combined*	*One Week Recall, Personal & Telephone*	*Scaling Method, First Wave Only*
Life	59.6%	58.2%	56.2%
Newsweek	25.2	25.4	27.5
Saturday Evening Post	45.0	42.0	47.0
Sports Illustrated	8.0	13.7	13.8
Time	36.6	33.3	28.7
U.S. News & World Report	14.3	17.0	12.6
(Base)	(238)	(481)	(87)

Such probing would not usually be practicable in a full scale study. The differences found in Table 9, then, are probably attributable to question design only.

TABLE 9

ANALYSIS OF VARIANCE OF ESTIMATES BY METHOD AND WAVE

Source of Variation	*Sum of Squares*	*Degrees of Freedom*	*Mean Square*	*F*
Magazines	8,909	5	1781.8	54.39*
Methods (Scaling *vs.* Recognition)	235	1	235.0	7.16*
Waves	11	2	5.5	<1
Interactions:				
Magazines—Methods	194	5	38.8	<1
Magazines—Waves	105	10	10.5	<1
Methods—Waves	320	2	160.0	4.87*
Magazines—Methods—Waves	326	10	32.6	
Total	10,100	35	——	

* Significant beyond the .05 level.

The focus of this paper is on interviewing and questioning methods. Having found the Scaling Method to state average readership much as do other methods, we must now test how consistent this relationship is.

It is relatively unimportant how large or small the method's fixed bias is in estimating various magazines' audiences. What is essential is that the method over-estimate or under-estimate all magazines' audiences to the same degree.

Note the significant difference between the Scaling Method and the Recognition Method by referring to Table 9. This analysis seems to establish that the Scaling Method estimates are consistent with the Recognition Method estimates, but subject to slightly higher experimental error.

Curiously, however, the high mean square for the Methods-Waves interaction suggests that the tendency for the Scaling Method to underestimate the Recognition Method in all three waves is largely an artifact in this experiment. This conclusion is based on the fact that the Scaling Method is significantly subject to the conditioning inherent in this experimental design, while the other methods are not. Notice in Table 10 how the Scaling Method estimates decline from wave to wave, and

TABLE 10

PER CENT CLAIMED READING, BY TWO METHODS

	Recognition	*Scaling*		
Magazine	*(All Waves)*	*Wave I*	*Wave II*	*Wave III*
Life	58.7%	56.2%	43.0%	44.3%
Newsweek	25.2	27.5	14.0	17.1
Saturday Evening Post	45.0	47.0	34.9	31.4
Sports Illustrated	8.0	13.8	10.4	10.0
Time	36.6	28.7	31.4	30.0
U.S. News & World Report	14.3	12.6	14.0	5.7
(Base)	(238)	(87)	(86)	(70)

that the first wave's estimates total almost the same as the Recognition Method estimates. We conclude, then, that the Scaling Method, if free of conditioning and based on a strict sample, can estimate magazine audiences reasonably well.

Reading Habits

A question on reading habits, as discussed previously, was asked in the following form:

About how often would you say you actually read———?

The last column of Table 2 shows the estimates. For all its theoretical defects and editing difficulties, we found generally good results from this question. However, this should probably be ascribed to careful probing in the interviews and editing of the responses. Such effort would probably not be practicable in a full scale study.

Mail Questionnaire

The mail questionnaire asked respondents to check their past reading of each magazine "within the past two weeks," "three to four weeks ago," "more than four weeks ago," and "do not read". Table 2 compares (in columns 3, 5 and 6) the three kinds of two week Recall estimates. Here again we must conclude that traditional objections, in this case to mail questionnaires, may not be too serious in practice.

REFERENCES

POLITZ, ALFRED, RESEARCH INC., *The Readers of the Saturday Evening Post (Technical Appendix).* Philadelphia: The Curtis Publishing Company, 1957.

STOCK, J. STEVENS. *Characteristics of the Reading Audiences of Newsweek, Time and U.S. News & World Report (Technical Appendix).* New York: Newsweek, 1959.

WHY THIS ISSUE IS NUMBER 5

Many forms of reference notation give a journal article's volume or year but not both. So that this *Journal's* future volumes will each correspond to a given year, Volume 1 will have six issues (September, 1960 through December, 1961). Thereafter each volume will contain four issues, Volume 2 those of 1962, Volume 3 those of 1963, and so on.

A Survey of Case History Advertising

GEORGE FISK
University of Pennsylvania

The case history is being used as copy by more and more advertisers. Professor Fisk reports who they are, why they use it, how much it costs them and what their problems are.

THE DOLLAR VOLUME of case history advertising and sales promotion may range from half to three-quarters of a billion dollars a year. This conclusion is based on the published advertising budgets of case history advertisers analyzed by their estimated allocation to case history advertising in the writer's survey.

Expenditures for case history advertising might rise appreciably above the billion dollar level if existing opportunities for case history promotion were fully exploited. Yet beyond the occasional trade paper success story there is little written to aid the sales and advertising executive to identify and appraise opportunities for employing this type of promotion. The writer's study was undertaken to answer the following questions:

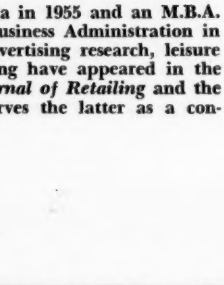
What kinds of firms now find case history promotion effective?

What kinds of information should case history advertising provide?

What are the common problems in collecting and preparing cases?

The term *case history* will be used in this paper to refer to an advertisement or sales presentation which relates the real life experiences of a customer who finds the product a satisfactory solution to his buying problem. It must be a case and it must be history. While the term *case history* will be used throughout in connection with advertising, it should be remembered that case histories are also used in connection with personal selling.

Case histories, for which material has been especially prepared, are distinguishable from certain paid testimonials. In the former, the sales or advertising message consists of a statement of a problem and a range of available alternatives analogous to those confronting the prospective customer. Testimonial advertising documented by proof of savings of time, cost, or effort, or by sales increases following purchase of the advertised product, are considered case history promotion. Testimonials or endorsements based on personal satisfaction with the product are not so considered.



GEORGE FISK is an associate professor of marketing in the Wharton School of Finance and Commerce at the University of Pennsylvania. He has studied advertising in and out of advertising agencies since 1945 when he was assistant research director at Sherman K. Ellis, now C. J. LaRoche. More recently he has been a consultant to the Bell Telephone Company of Pennsylvania, Alderson Associates, Sidney Hollander, Jr. and N. W. Ayer. Professor Fisk received a Ph.D. in economics from the University of Pennsylvania in 1955 and an M.B.A. from NYU's Graduate School of Business Administration in 1949. His papers on marketing, advertising research, leisure spending-behavior and city planning have appeared in the *Public Opinion Quarterly*, the *Journal of Retailing* and the *Journal of Marketing*. He also serves the latter as a consulting editor and reviewer.

USERS

Many inferences in this appraisal are based on replies provided by 192 respondents to a mail survey addressed to 396 firms initially believed to be case history users according to the definition above. The list of 396 firms was compiled by selecting the names of firms which qualified according to the definition which was used in a content analysis of national print media. Twenty of the 396 firms contacted indicated that they had little or no experience with case history advertising, so that the 192 firms and divisions of firms which did reply represented 51 per cent of the remaining 376 contacted.

Since nothing is known about non-respondents' case history promotion, the statistics to be presented can represent only a fraction of the population of case history users. For this reason, no inferences are made concerning this larger group. However, inasmuch as firms replying to the questionnaire collectively spent nearly a billion dollars for advertising in 1959 according to published sources, their replies are perhaps more representative of large firms than of small.

The typical case history user in the sample is a large firm selling manufactured products to industry. A majority (54 per cent) appear to be businesses with sales of $50 million or more per year, while only a small fraction (4 per cent) reported annual sales of under $5 million. Fifty-three per cent sold to industrial users exclusively, 42 per cent sold in both "ultimate" and "industrial user" markets, while only five per cent sold to "ultimate consumers" only. Sixty-seven per cent of the respondent businesses sold tangible products only, 25 per cent sold products and services, and eight per cent sold services only.

A relatively small proportion of case history advertisers used this method for most of their advertising, although 36 per cent estimated that more than half of their advertisements were case histories. However, 39 per cent estimated that between a half and a tenth of their ads were of the case history type. Altogether, three-fourths of all the firms responding estimated that 10 per cent or more of their ads were case histories. For the population of case history users, the true proportions using 10 per cent or more of their ads in case history form would probably be somewhat below three-fourths, due to the well-known tendency of the most interested persons to respond more frequently to mail surveys than is true of less interested persons included in the sample design.

USES

Case history advertising is most useful in presenting documentation of rational reason-why appeals for which a high degree of believability is required. This combination of requirements is most often found in promotion aimed at businessmen although it has also proven useful in the purchase of certain consumer durables and installation goods. Whenever the solution to a buying problem can be solved by analogy, despite widely differing levels of buying information, or whenever a number of persons who participate in a single buying decision must be appealed to by rational reason-why copy, there exist excellent opportunities for the use of case history promotion.

Believability

Gaining an increase in believability is among the most important reasons for using case history promotion. About four-tenths of the advertisers planning to increase case history advertising in the next five years cited gains in believability and conviction as a reason, and more than a quarter asserted that they had evidence that cases were effective in persuading people to buy. Much, if not most, of the increase in believability achieved through the use of case histories stems from the ease with which the prospect can place himself in an analogous situation.

However, in a number of instances, this gain arises because additional investment can be used to produce local cases. These appeal to the prospect who knows personally individuals employed by firms which have solved problems analogous to his—partly because he can personally validate the evidence, and partly because of his rivalry with persons he knows to be good businessmen or intelligent consumers.

There is a danger that case histories may be only a temporary solution for the problems of an advertiser who has otherwise overstated his claims, and who has lost the right to be believed. In the words of an oil company executive:

> It [the case history] might be construed as a rebuke to the advertising profession, indicating that the claims of an advertiser are no longer as believable as the statements of a user.

Versatility

At the time of the study in 1959, a number of the respondents had made, or were planning to make, changes in the ways case history information

is used. As shown in Table 1, publicity, direct mail promotion and help to the salesmen were among the most frequently mentioned uses, although a variety of other uses also received mention.

TABLE 1

NUMBERS OF RESPONDENTS WHO MADE OR PLANNED CHANGES IN USE OF CASE HISTORY INFORMATION

Method of Use	*Changes Made*	*Changes Planned*
Publicity. External house organ with wide circulation.	15	4
Direct mail. Heavy emphasis on direct mail; broadside with localized case history.	14	9
Use of direct mail to prepare way for salesmen, or to use as a sales tool.	12	6
Printed case history book for use by salesmen.	9	5
Furnished salesmen and/or publicity people with pictures and stories before ads appeared.	5	1
As means of communication to distributors and sales force.	3	2
News release.	3	1
For sales training (via external house organ).	3	—
Special mailing of reprints to problem areas.	3	1
Reprints as bill enclosures.	2	1
Other	3	5
No answer	16	11
Total	88	46

The experience of the Superior Tube Company, Collegeville, Pennsylvania, illustrates how a firm can take advantage of the versatility of cases. This firm developed a case based on the use of its 0.052 inch diameter tubing for the center arbors to hold the sweep second hand pinions of watches. The story, an excellent illustration of the use of precision tubing, was released to selected trade papers in the design, metal, purchasing, machining and jewelry industries. Interest in the end product—watches—also made the case suitable for ads in media such as *Business Week* and *Fortune*. Company salesmen found reprints such a good talking point that mailings were made up to be sent to Superior's 2,500 customers and key prospects to stimulate demand, and to the firm's distributor salesmen to win their support for the current public relations effort in the trade press. The same case material was found useful in training courses and sales meetings held at distributors' warehouses. Thus, costs on a per application basis were quite low and the company was able to use the same case for six different promotional purposes.

Economy

Among firms surveyed to determine total costs of preparing cases, median total costs per advertisement were in the vicinity of $200-$300 for collecting information and between $400-$500 for the preparation of an advertisement utilizing the information collected. Median costs for all promotional uses other than advertising were about $300 per case. Excluding media costs, expenses for all promotional uses of a single case were in the vicinity of $1,200, although as shown in Table 2, a small minority estimated average total costs to exceed $2,000 per case.

TABLE 2

RESPONDENTS GROUPED BY COSTS OF PREPARING THEIR CASE HISTORIES

Estimated Average Cost	*Collection of Information*	*Preparation of Advertisement*	*Total Investment Other Uses*
Under $100	14	2	11
100- 199	37	10	8
200- 299	32	17	21
300- 399	22	13	10
400- 499	9	25	4
500- 749	11	36	8
750- 999	4	13	5
1,000-1,999	1	9	15
2,000-2,999	—	4	2
3,000-4,999	—	2	—
5,000 & over	—	3	3
Not available	29	20	28
Varies	9	12	16
No answer	24	26	61
Total	192	192	192

The experience of a Philadelphia insurance firm illustrates why case history preparation costs can be relatively modest. This firm found it could use claim records to illustrate a variety of convincing situations for which insurance had previously provided a means of dealing with unanticipated and uncommon risks. The only preparation costs were those incurred for producing photographs of the circumstances leading to the filing of claims. No other approach used by this company has produced such consistently low average costs per inquiry when all costs of space and preparation were considered.

Effective Communication

Any good advertisement is designed to achieve a number of communication objectives: comprehension, believability, recall, persuasion and a high

correlation between audience penetration and subsequent purchase of the proffered service or commodity. Nevertheless, assuming that the audience has been initially selected by the choice of media, headline, etc., and assuming further that the product is one for which rational appeals can be communicated by cases, the case history method has demonstrated to survey respondents its ability to achieve each of these objectives.

One advertising manager for an industrial supply product declared flatly that the case method produces more traceable sales than any other method, and for this reason the bulk of his company's advertising employs case histories. While this statement assumes that advertising works by "hypodermic effect" rather than by serving as one of a number of predisposing stimuli associated with any given level of sales, it does convey the conviction born of experience with the promotional effectiveness of cases.

One hundred fifteen respondents, or 60 per cent of those who replied, said they had evidence that case history advertising had been more effective for their company's purposes than other advertising methods, 19 per cent estimated that it was about the same in effectiveness, and the remainder found it less effective or had no evidence. Among the latter group were a number who questioned all methods of measuring advertising effectiveness.

Among those who accepted some form of evidence, 50 respondents or 26 per cent of the 192 respondents mentioned inquiries, unsolicited comments from readers, requests for reprints or approval of the sales force as reasons for believing cases to be more effective than other advertising techniques. Readership studies were accepted as evidence by 32 per cent. One particularly interesting belief was that in trade circles case histories of large companies influenced others in the use of a firm's products, and also tied up the sales of the big company for a long time.

Obviously caution must be exercised in accepting as evidence for all firms criteria of success that may be suitable for only one firm. The relationship between advertising and sales, particularly for products with a long life and a low replacement rate, is tenuous at best. The construction of quantitative mathematical sales response models, even for products with a high replacement rate, often involves assumptions too heroic to gain ready acceptance from seasoned advertisers. Criteria of success obviously depend on the advertising objectives of the company, and while sales response to advertising is often unmeasurable, the more limited objectives of advertising are measured with ease providing only that the company knows what it wants its advertising to do.

Contrasting Examples

For example, it is fairly simple to measure the association in the public mind between Caterpillar tractors and the road repairs essential to reduction of hazardous driving conditions. It is also relatively easy to measure readership, comprehension or "reader impression" and believability of Caterpillar ads which describe road hazards which could be eliminated by proper highway maintenance and improvement programs. By these standards of evidence, the Caterpillar Company has developed an outstanding image of a firm concerned with the outmoded highway system—a problem heavily freighted with public welfare considerations.

In contrast, the efficiency of U. S. Plywood's case history advertising for Weldwood marine plywood relates to the rate of increase in orders by pleasure boat builders and purchasers. Judging from published sources, "effectiveness" can be defined as the slope of the traceable sales response curve for all case history ads as compared to the slope of the curve generated by other advertising and sales promotion methods. While the validity of such a measure may be questioned because of joint effects, lags between exposure to advertising and the sales response and so forth, the concept of an immediate response rate (within 30 days) is not ambiguous. Moreover, since the company's stated advertising objective is to inform, it can study before and after information levels with respect to: (1) familiarity of boat builders with Weldwood marine plywood construction economies, (2) familiarity of boat buyers with durability and maintenance characteristics, and (3) familiarity of both builders and buyers with the strength and stress properties under high speed boating. By assumption, information levels are related to sales, but this does not preclude measurement of immediate traceable sales response.

In summary, then, the advantages of case history promotion are a high degree of believability, versatility in numerous uses, economy and amenity to traceable sales and inquiry measurement. Now let us assess its major limitations as a promotional tool.

LIMITATIONS

Case history promotion appears unsuitable for the following products:

1. Consumer products requiring emotional rather than rational appeals.
2. Scientific or technical products for which exact specifications must be known before the prospect can make a buying decision.
3. Products with geographically thin markets or relatively few buyers.
4. Products for which the use produces highly variable results depending on operating conditions or environment.
5. Products for which the use results in competitive cost savings or which are used in secret production processes for which case history data are hard to obtain.

With respect to the first limitation, case histories are not testimonials with emotional appeal, but rational solutions for common problems. How, then, could an acceptable case history be written for a perfume? In consumer advertising, case history applications are largely of the long-range planning or deliberate purchase variety. Headlines such as "How We Redecorated Our Home in Just One Day," or "How We Retired in Fifteen Years on $300 a Month" are typical of these case history ads. Emotional appeals involving visceral drives are seldom in evidence in case history advertising, and for this reason the method is not used in consumer advertising as much as in business promotion.

Second, the exacting demands of scientists and technicians for precise product specifications restrict the usefulness of cases because in them problems must be stated, alternatives in the form of present and proposed practices explained, and the superiority of the solution employing the promoted product must be demonstrated. Although common engineering problems of a recurring nature can often use cases side by side with product specifications, highly technical applications of a product are usually unknown to anyone but to the individual scientist whose problem is nearly always unique. Under these circumstances, much of the information in cases is irritatingly irrelevant and the prospect is repelled rather than attracted by the information presented. Even among the 192 respondent firms that use cases, the problem most frequently mentioned is difficulty in locating users and usages of the product or services which have unusual or wide appeal.

The experience of the truck and coach division of an automotive manufacturer illustrates the common situation in which several shortcomings occur jointly. Not only were prospects with similar truck operating problems too few to warrant the development of cases, but fuel economy, a major selling point, varied so widely from one set of operating conditions to another that cases with widespread applicability were difficult to develop.

A major American airline experienced great difficulty in finding case histories of nationally recognized firms which might be readily identified by prospective users of air freight service. Sales and inquiries promoted by the cases developed were so disappointing that the case history approach to air freight was dropped.

The final limitation—difficulty in obtaining case histories—is in fact a complex of several lesser shortcomings. In highly competitive industries there is often a rapid rate of technological development, and under these circumstances it is difficult to develop a backlog of cases before they become obsolete. Costs and productivity have risen rapidly in recent years. The discovery of substitute materials or processes is so swift and secret that a case approach can seldom provide information rapidly enough. It is often impossible to recruit users of a product to endorse it if the endorsement means revealing trade secrets. This limitation applies to firms which do use case history promotion as well as to those which do not.

COLLECTING CASES

Assuming that the foregoing limitations reveal no major obstacle to the use of cases, a method of collection should be established. The experiences of firms already using case histories can help others pick a suitable procedure.

Salesmen's calls on customers are used to some degree by a majority (78 per cent) of the firms surveyed, and field trips by members of the advertising department were mentioned by 67 per cent of the respondents. A smaller proportion, 41 per cent, mentioned advertising agencies; reporter-photographer agencies were used by 38 per cent, and mail questionnaires to users of the company's product were mentioned by 17 per cent.

Problems encountered in collecting case histories, with the proportion of survey respondents who mentioned each, were as follows:

Locating users and usages of the product with unusual or widespread appeal (43 per cent).

Unwillingness of user to endorse product for a variety of reasons ranging from industrial secrecy to public relations with other suppliers (22 per cent).

Difficulty in obtaining good photographs and documentation, especially from users located at a distance (10 per cent).

Lack of sufficient time and personnel to do the job (7 per cent).

Difficulty in timing due either to obsolescence or timing of request to coincide with inauguration of operation on new equipment in customer's plant (4 per cent).

Selecting the best or most representative case histories for the entire country from a number of possible choices (3 per cent).

Since the solutions proposed to solve each of these problems have much in common, they may be considered jointly with but a few exceptions. For example, to locate users and usages of wide appeal and to secure endorsements by reluctant users of the product, salesmen must be trained to recognize the type of case history which can benefit both them and their company, so that they may obtain endorsements as part of the negotiation of the transaction. While the salesman is locating product usages with wide appeals, he may also initiate preliminary contact with top level executives or advertising people in the prospect company. During the sales interview, he may show how advertised use of the product by the prospect will give the prospect advertising and public relations benefits he would not otherwise obtain. In return for his extra work in finding and developing good case histories the salesman must often be compensated by credit toward his sales quota, by prizes for the best case submitted or by other incentives furnished in individual sales compensation plans.

When customers are unwilling to provide salesmen with endorsements, statistics on cost savings or productivity gains, some firms let their advertising departments or case history writers follow leads furnished by salesmen. By having someone other than the salesman explain the benefits and by encouraging a free exchange of data, background materials and statistical information can often be secured from otherwise reluctant customers.

The next two problems may be dealt with jointly. Lack of personnel to do the job properly can be overcome by employing an outside specialist agency to prepare both the photography and the copy for cases involving firms located at a distance from the case history advertiser. As in other types of work for which personnel needs arise at sporadic intervals, the costs of outside specialists are often lower than the costs of the same work performed internally. Photo-reporter services and advertising agencies are two types of outside specialists qualified for this work. To maintain the quality of case histories a number of companies have developed detailed instruction sheets for photographers, reporters and advertising agencies. Since what is everybody's business is nobody's business, it is sometimes desirable to assign the responsibility for supervision and coordination of the work of the outside specialist agency to the advertising agency account executive or to his counterpart within the advertiser's own organization.

The last two problems deal with matters of selection criteria. If case histories must be up-to-date, the danger of obsolescence of backlog material can be minimized by maintaining a tickler file system with arbitrary cut-off dates for any cases which initial screening indicates to be interesting but perishable. If salesmen or product managers are themselves responsible for using case material before it becomes obsolete, the marketing administrator gains both in fuller utilization of cases and more effective selling at the same time. Occasionally, too rapid a rate of obsolescence indicates that the case history space budget is inadequate and that more frequent use of fresh material or more insertions of currently used material is called for.

In selecting the best from among many good case possibilities, a check list based on experience has been developed by some firms, but as in most promotion there can be no substitute for individual pretesting of comprehension, believability and relevance to the problems confronting the customer. The survey was noteworthy for the inarticulateness of respondents on this subject. Although pretesting was not suggested to respondents as a method for guiding the selection of cases, their failure to mention its use may be indicative of an under-utilized technique for improving the effectiveness of case history advertising.

CONCLUSION

Nobody knows exactly how much is being spent for case history advertising and promotion, when it is more effective than other forms of persuasive communication, for which specific products it is best suited, and for whom in particular case history promotion has the most appeal. Nevertheless, as the modest survey presented in the preceding paragraphs indicates, there are substantial opportunities for its use. One conclusion comes close to being indisputable: approaching opportunities to use case history promotion appear much smaller from before than chances missed look from behind.

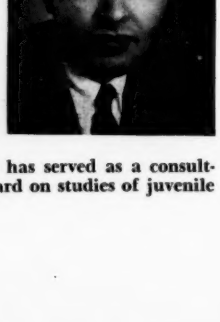
Selecting Commercial Spokesmen

ARTHUR J. KOVER[1] and SEYMOUR LIEBERMAN
Kenyon & Eckhardt Inc.

How do viewers rate a TV spokesman they like? In this study, as believable, sincere and down-to-earth—but definitely not cold and dull.

THE SELECTION of a spokesman for a product or product class is a recurring and difficult problem for both advertising agencies and their clients. Large amounts of money and energy are often spent to build an association between the product and the spokesman. Examples are Vaughn Monroe for RCA, Betty Furness for Westinghouse, and John Cameron Swayze for Timex.

ARTHUR J. KOVER is a project supervisor in the research department of Kenyon & Eckhardt. Before coming to that agency in 1960 he was assistant supervisor of the Study Skills Office at Yale University; there he investigated social and psychological factors affecting academic performance. He received a B.A. from Cornell University in 1953 and an M.A. from Yale in 1954. He is a member of the American Sociological Association, the Society for Applied Anthropology, and the Medieval Academy of America.



SEYMOUR LIEBERMAN is a vice president and marketing group head at Kenyon & Eckhardt. Previously he was an associate research director in the agency's research department. He received a B.A. in 1948 from Brooklyn College and an M.A. in 1949 and a Ph.D. in 1955 in social psychology from the University of Michigan. From 1948 to 1955 he was a study director in the Survey Research Center at the University of Michigan, where he conducted research on the structure and function of large organizations. He has served as a consultant to the New York City Youth Board on studies of juvenile delinquency.

Through his attention-drawing power or the effect of his personality, an effective spokesman may add significantly to the impact of a product's advertising campaign. On the other hand, an ineffective spokesman (e.g., one who is somehow inappropriate for the product which he represents) may actually detract from the impact of the product's campaign. Or a spokesman may neither add to nor detract from the effectiveness of a product's campaign, in which case it would not have been worth the additional money and energy expended to secure his services.

This study was designed to answer one question about effective spokesmen: which personality traits best distinguish a spokesman who is liked from one who is not?

It should be acknowledged that the effectiveness of a spokesman should be evaluated not in terms of his capacity to induce consumers to like *him*, but rather in terms of his ability to induce consumers to like the *product* advertised. However, liking a spokesman may be construed as one criterion of his effectiveness because, other conditions being equal, a well-liked spokesman is apt to do a better job of increasing consumer acceptance of the advertised product.

While the product in this study was cars, the general approach may be applied to other product classes, as well as to other criteria than being well-liked and to the general problem of effectiveness itself.

[1] Now at Ogilvy, Benson & Mather.

METHOD AND RESULTS

Personal interviews with 590 men in car-owning households were conducted in six markets selected to provide variation in geographical region and size of market.

The total sample was randomly divided into two equal groups. One group was shown a television commercial for a make of car in which Spokesman A was featured. The other group was shown a television commercial for another make of car in which Spokesman B was featured. The commercials were shown by means of a portable movie projector which interviewers brought into respondents' homes.

A and B were selected to represent different types of spokesmen. They differed in age, physical appearance and personal mannerisms. They presented a range of personal traits which may or may not have been appropriate for automobile spokesmen.

Respondents exposed to each spokesman were asked to rate their liking of the spokesman on a five-point verbal scale which ranged from "like very much" to "dislike very much."

Table 1 shows respondents' ratings of the spokesmen to whom they were exposed. Both men were rather well-liked, although more respondents reacted favorably towards Spokesman A than toward Spokesman B.

TABLE 1

LIKING OF THE SPOKESMEN*

	Spokesman A	*Spokesman B*
Liked him very much	36%	31%
Liked him fairly much	41	35
Neutral	19	31
Disliked him somewhat	4	2
Disliked him very much	0	1
TOTAL	100%	100%
BASE	298	292

* Based on the question: "How do you feel about the person who gave the commercial—would you say you liked him very much, liked him fairly much, felt neutral about him, disliked him somewhat, or disliked him very much?"

Besides rating their liking of the spokesman, respondents also rated him from zero to ten on various personality traits. The more respondents saw a trait as fitting the spokesman, the higher the number they were to give him on that trait; the less respondents saw a trait as fitting a spokesman, the lower the number they were to give him. (In another study, the list of traits might be different depending on the specific problem and the product or product class under consideration.)

Table 2 shows the ratings given the two spokesmen for each of the traits studied. Both spokesmen tended to receive relatively high scores for positive traits and rather low scores for negative traits. This supports the finding in the previous table that both spokesmen were generally well-liked.

TABLE 2

MEAN PERSONALITY TRAITS OF THE SPOKESMEN*

Personality Trait	*Spokesman A*	*Spokesman B*
Mature	8.7	7.8
Intelligent	8.4	8.1
Sincere	8.1	7.8
Friendly	8.0	7.7
Relaxed	7.9	7.4
Believable	7.8	7.3
Down-to-earth	8.0	7.1
Manly	7.6	7.5
Pleasant-looking	7.5	7.7
Strong	7.6	7.5
Cultured	7.7	7.3
Lively	7.0	7.3
Good sense of humor	7.4	6.6
Forceful	6.3	6.5
Youthful	4.3	6.1
Stiff	2.7	2.2
Cold	1.9	2.0
Out-of-date	2.3	1.7
Dull	1.9	1.8
Weak	2.2	1.8
BASE	298	292

* Based on the question: "Now I'd like you to use this scale (0 to 10) to tell me how well various descriptions fit the person who just gave the commercial. The more a description fits him, the bigger the number you should give him. The less a description fits him, the smaller the number you should give him. Now, how well does this description fit him? . . ."

But there were also differences in the ways the two spokesmen were perceived by respondents. Spokesman A was seen as more mature, more down-to-earth, and more as having a sense of humor than Spokesman B. Spokesman B was seen as more youthful and less out-of-date than Spokesman A.

The next step in the analysis was to determine the extent to which each of the personality traits was associated with liking of the spokesman. Respondents were divided into two groups: those who liked the spokesman (defined as those who checked "like very much" and "like fairly much") respondents who did not like the spokesman (defined as those who checked "neutral," "dislike somewhat," and "dislike very much"). Ratings on each trait were computed separately for those who did and for those who did not like the spokesman.

For each trait, the difference between the rating given by those who liked him and those who did not like him was termed a "difference score." The larger this difference score for any trait, the more that trait differentiated those who liked the spokesman from those who did not. In other words, the larger the difference score, the more important that trait is in distinguishing between "liked"

TABLE 3

RELATIONSHIPS BETWEEN LIKING AND PERSONALITY TRAITS OF SPOKESMAN

	Spokesman A			*Spokesman B*		
Personality Trait	*Liked Spokesman*	*Did Not Like Spokesman*	*Difference Score*	*Liked Spokesman*	*Did Not Like Spokesman*	*Difference Score*
Mature	8.8	8.2	+0.6	8.2	7.1	+1.1
Intelligent	8.6	7.8	+0.8	8.6	7.1	+1.5
Sincere	8.9	6.5	+2.4	8.5	6.5	+2.0
Friendly	8.7	7.1	+1.6	8.1	7.0	+1.1
Relaxed	8.2	6.6	+1.6	8.0	6.1	+1.9
Believable	8.2	6.4	+1.8	8.1	5.7	+2.4
Down-to-earth	8.4	6.6	+1.8	7.8	5.6	+2.2
Manly	7.8	6.7	+1.1	7.9	6.8	+1.1
Pleasant-looking	7.9	6.3	+1.6	8.2	6.7	+1.5
Strong	7.9	6.5	+1.4	8.0	6.3	+1.7
Cultured	8.0	6.5	+1.5	7.9	6.1	+1.8
Lively	7.4	6.0	+1.4	7.6	6.6	+1.0
Good sense of humor	7.8	5.8	+2.0	7.3	5.3	+2.0
Forceful	6.5	5.7	+0.8	6.7	5.9	+0.8
Youthful	4.6	3.6	+1.0	6.5	5.4	+1.1
Stiff	2.5	3.5	—1.0	1.8	3.0	—1.2
Cold	1.5	3.4	—1.9	1.7	2.8	—1.1
Out-of-date	2.1	3.3	—1.2	1.3	2.4	—1.1
Dull	1.5	3.5	—2.0	1.2	2.9	—1.7
Weak	1.9	3.2	—1.3	1.5	2.4	—0.9
BASE	232	66		195	97	

spokesmen and "not liked" spokesmen for the product or product class under consideration.

Table 3 shows the difference scores for each spokesman. As might be expected, those who liked each spokesman gave him higher scores on positive traits than did those who did not like him, while those who did not like the spokesman attributed more negative traits to him than did those who liked him.

This gives us information about the most important character traits for *each* spokesman. However, this is not yet the information we were seeking in this study. Effectively, Spokesmen A and B were considered as a sample of possible spokesmen for cars. The information we wished to derive from these difference scores was the traits which *generally* characterize well-liked spokesmen for cars.

In order to obtain this more general information, some idea should first be obtained of the extent to which the two distributions of difference scores are similar. To determine if the rank distributions of difference scores for groups exposed to Spokesman A and Spokesman B were effectively similar, a rank difference correlation coefficient was computed for the two distributions. Using the correction for tied scores (Siegel, 1956), r_s equaled + .63. Thus larger difference scores for Spokesman A also tended to be larger difference scores for Spokesman B, showing that traits associated with liking A were also associated with liking B.

In computing the correlation coefficient, signs of the difference scores were disregarded since we were interested in determining the extent to which each trait was associated with liking of the spokesman. Had the signs not been disregarded, this would have produced a spuriously high correlation coefficient. The signs were used, however, in determining whether a given trait was to be sought or avoided in a spokesman.

An even higher correlation between the difference scores for the two spokesmen might have been obtained had they delivered commercials for the same car rather than for different cars. For although certain traits might tend to distinguish well-liked spokesmen for the over-all product class of cars, some traits may be more relevant for one car than for another car. The extent to which a trait is relevant may depend on the price class of a car or on the particular kind of image which has been built up for it.

But despite the fact that the two spokesmen delivered commercials for two different cars, there was still a fairly high correlation between the "difference scores" for the two spokesmen. This finding gives some assurance that these results can be generalized; if the ranking of "difference scores" was so alike for two such dissimilar spokesmen, it is possible that the findings from this two-man sample of spokesmen may have some use in arriving at a fairly general definition of the traits needed in well-liked automobile spokesmen.

The last step was to select those traits with the highest absolute difference scores for both spokesmen combined. Table 4 shows that of the positive

traits, "sincere," "believable," "down-to-earth," and "good sense of humor" best distinguished "liked" from "not liked" spokesmen. Thus these traits should be looked for first in selecting a well-liked spokesman for cars. "Cold" and "dull" were the negative traits which best distinguished "liked" from "not liked" spokesmen, and should be avoided in selecting a well-liked spokesman.

TABLE 4

MEAN DIFFERENCE SCORES FOR PERSONALITY TRAITS OF SPOKESMEN A & B COMBINED

Sincere	+2.2	Intelligent	+1.2
Believable	+2.1	Manly	+1.1
Down-to-earth	+2.0	Youthful	+1.1
Good sense of humor	+2.0	Mature	+0.9
		Forceful	+0.8
Relaxed	+1.8		
Cultured	+1.7	Stiff	—1.1
Pleasant-looking	+1.6	Weak	—1.1
Strong	+1.6	Out-of-date	—1.2
Friendly	+1.4	Cold	—1.5
Lively	+1.2	Dull	—1.9
BASE	590		

DISCUSSION

The above findings should be limited to automobile spokesmen. Traits which characterize a well-liked spokesman for one product class may differ from those of a well-liked spokesman for some other product class.

For example, Table 4 shows that believability was more important than friendliness as a characteristic of well-liked automobile spokesmen. While this finding might hold for the automobile field, it will not necessarily hold for other product fields. The relative importance of believability and friendliness might vary depending on the particular product category under consideration. Conflicting claims made about different makes of automobiles, the none-too-flattering image consumers have of automobile salesmen, and the large financial obligation incurred when a car is selected probably place a premium on the credibility of the automobile spokesman. Under these conditions, consumers may have greater acceptance of a spokesman who provides the reassurance of being honest and straightforward than one who simply plays the role of the friendly person.

For another product, say children's breakfast cereal, the reverse might be true. Whether a spokesman is sincere or insincere may not enter children's minds, but whether the spokesman is a warm, accepting adult or a hostile, threatening authority figure may be of central concern. Here friendliness would be of greater importance than believability in determining liking of the spokesman.

It might also be noted that traits which are relatively unimportant in distinguishing "liked" from "not liked" automobile spokesmen could still be important in differentiating spokesmen for other product fields. For instance, Table 4 indicated that neither of the two traits having to do with age—"mature" and "youthful"—was particularly efficient in distinguishing "liked" from "not liked" automobile spokesmen. But the age of the spokesman for other products may be quite relevant to whether consumers accept him. This is more likely when the age of the consumer is related to his use of the product, or when the consumer's age is related to the functional or psychological significance of the product. A laxative may require a somewhat older spokesman because it is associated with the problems of advancing age; a cosmetic may call for a younger spokesman (or spokeswoman) because it is associated with youth.

Much of this discussion is necessarily conjecture and speculation. Our intention is to suggest that techniques similar to those described here should reveal more about the patterns of traits which characterize effective spokesmen for different product categories, to specify the conditions under which a given trait will or will not be relevant, and to develop rules and principles for matching product and spokesman more efficiently.

SUMMARY AND CONCLUSIONS

This paper described a method for determining which personality traits are most closely associated with liking of a spokesman for a given product or product class. After exposing different spokesmen who delivered a commercial to different but matched groups of respondents, we measured the latter's liking of each spokesman and their ratings of each spokesman on certain personality traits. Analysis revealed those traits which best distinguished liked from not liked spokesmen. "Believable," "sincere," "down-to-earth," and "good sense of humor" were positive traits to be sought in a well-liked spokesman for automobiles; "cold" and "dull" were negative traits to be avoided.

We pointed out that these traits may not be important in spokesmen for other products. We suggested that future research using similar techniques could develop principles which would help select more effective commercial spokesmen.

REFERENCE

SIEGEL, SIDNEY. *Non-Parametric Statistics for the Behavioral Sciences.* New York: McGraw-Hill Book Company, 1956.

County Business Patterns

INGRID C. KILDEGAARD
ARF Research Statistician

Often advertising to industry is like shooting in the dark—at thousands of moving targets. Miss Kildegaard describes a set of reports, published this month, which sheds needed light on these targets.

INDUSTRIAL advertisers require yardsticks quite different from those used in consumer advertising. Their markets—industries and institutions rather than individuals and families—cannot be described by distributions of population, age and income. Sometimes they can barely be described at all.

Each industry has its own geographic distribution and may be concentrated in a few areas. But the industrial advertiser's total market is usually made up of many industries. To define this market he needs to identify each industry in it, and determine its relative importance.

His surveys may show that an industry's buying decision is a complicated process with many individuals participating. He may find that successful copy for one industry group has no appeal for another. Similarly, media to one group may by-pass another. And competitive activity may vary by geographic area as well as by industry.

No single system can solve all these problems. But if the industrial advertiser has identified his market and rated its components, he can better measure sales potentials, allocate his advertising budget, and evaluate sales and advertising performance. Here the report, *County Business Patterns,* is a most important tool.

42 Million Employees

The *1959 County Business Patterns* is the most comprehensive statistical report of kind-of-business information available. The data are based on the first-quarter 1959 tax returns filed by employers under the Old-Age and Survivors Insurance program. Covered are 3,302,523 industrial and commercial establishments with 41,903,940 employees.

This set of reports comes in ten parts (15 books), showing statistics for the United States as a whole and for each of the nine Census geographic divisions. State and county kind-of-business details are given under the following broad industry divisions: agricultural service, forestry-fisheries, mining, contract construction, manufacturing, public utility,

wholesale trade, retail trade, finance-insurance-real estate, and services.

Shown for each industry or industry group are the number of business establishments reporting their taxable payroll for January through March 1959. The total number of employees is also reported and establishments are classified by number of employees.

Ninth Edition Enlarged

The current edition is the ninth in the series of *County Business Patterns*. They were first published in 1946 at the request of manufacturers of office equipment. Previously this information had been buried in the administrative files of the Bureau of Old-Age and Survivors Insurance. The first three editions were based solely on the BOASI returns; now the data are linked to the results of the Census of Business, Manufacturing and Minerals and are supplemented by a special survey of multi-unit establishments. Before 1956, data were given only for large counties; now virtually all counties are reported separately. Exceptions are New York City, where the five boroughs are combined, and the eight states with more than 99 counties. For these states smaller contiguous counties are grouped, since the tabulating system provides for a maximum of 100 reporting areas per state.

Two new features are included in the 1959 report. Many non-profit organizations are included for the first time. The Social Security Act was recently extended to permit their coverage on a voluntary basis. Second, there is a substantial increase in the amount of industry detail. Data are shown for every industry or industry group having at least 100 employees or ten reporting units.

SIC Groups

Reporting units are generally single establishments, classified according to their major activity. Industry data follow the definition given in the Standard Industrial Classification (SIC) Manual 1957 edition. There are 71 major industry groups each of which is assigned a two-digit code. Each is subdivided into component groups, one more digit being added to the code with each division. The manufacturer of pharmaceutical preparations, for example, is assigned a four-digit number 2834. This is a component of the industry group number 283, which includes all drug manufacturers. This group in turn belongs to the major industry group number 28, the classification for those industries manufacturing chemical and allied products.

The *1959 County Business Patterns* gives industry data for nearly all the three- and four-digit industries. In addition to the 71 major industry groups, the data are given for 350 three-digit industry groups and 596 four-digit industries. By contrast, the previous edition published information for only 125 three-digit industry groups and just 11 four-digit industries.

Market Analysis Tools

The Office of Distribution of the Business and Defense Services Administration has issued a booklet describing the application of these county data in marketing and advertising research: *Market Analysis Tools—County Business Patterns.* Five case histories show how these data might be used. The discussion is mainly directed to industrial marketing, but the interests of the consumer advertiser are also considered. As the booklet states, no statistics are perfect and the County Business data have their limitations. The data are arranged for all-purpose use and must be re-grouped according to the individual user's needs. Manufacturers, for example, are classified as though they produced one single line of goods while in fact they may have a diversified line. A large manufacturer's production of a secondary product, though more important than that of smaller establishments who manufacture only that product, may not be reported.

In spite of such defects, these county data have been used successfully in charting sales and advertising programs. They have proved especially valuable when coordinated with independent industry statistics.

The *1959 County Business Patterns* is being issued this month, September, 1961. To receive order blanks and price lists for part or all of the 1959 edition, write the Superintendent of Documents, U. S. Government Printing Office, Washington 25, D. C.

Factor Analysis

GWYN COLLINS
ARF Research Mathematician

A popular mystery in advertising research is here revealed as nothing but the attempt to reduce many measurements to few. Mr. Collins reviews the different methods and the disagreements among their users.

Factor analysis has become increasingly popular among advertising and market researchers. Its techniques and its terminology are, however, better known than the intentions which underly its purely mechanical procedures. Until quite recently, to become acquainted with factor analysis called for extensive reading of original papers, most of them scattered through the psychological journals. The publication of *Modern Factor Analysis* (Harmon, 1960), has changed this state of affairs considerably. This book brings together a large number of recent developments and ideas which are to be found in no other single source; at this time it is the most authoritative work on the subject. It is appropriate, therefore, to consider not just how to factor analyze, which this book explains with clarity and in detail, but the general intention of the factor analyst and, in broad conceptual terms, how his results are obtained and presented.

This article, however, will not discuss where and where not to use factor analysis. The general purposes of factor analysis have been stated as:

> . . . to reduce the number of experiments the scientist must do to impose order on complex subject matter. It helps him (or his manager) decide which experiments should be done next. So many responses are caused by large numbers of stimuli that his factor analysis may be the only way the researcher can reach a conclusion—within his lifetime—about which stimulus variables are most important in producing the response he wishes to predict or control. (Ramond, 1961.)

This topic will not be elaborated further here.

Briefly factor analysis is a technique for representing a large number of measurements each made on many objects or persons in terms of some smaller number of artificial measurements. For instance, the results of twenty different tests of mechanical ability may be factor analyzed and from the analysis we might find that from just four indexes, or factors, for each person tested, we could reproduce all twenty scores with a high degree of accuracy. This kind of thing is the basis of factor analysis. Unfortunately for simplicity, there are many ways of doing it and something needs to be said about them and their points of difference.

Factor analysis is the collective name given to all these procedures for deriving factors from original measurements. Its central idea was suggested 75 years ago as a means of isolating the fundamental characteristics of convicted criminals; its techniques were first introduced almost 60 years ago by a British engineer, Charles Spearman, in an article on intelligence in a psychological journal. Since then, its practice has been almost entirely confined to psychology, especially to the field of personality measurement. Spearman's paper was itself entitled "General Intelligence, Objectively Determined and Measured." It presented the technique as a means of resolving disputes about the meaning of intelligence, but the method he presented, and later elaborations of it, have served to feed controversy in this very area right up to the present time.

Professional mathematicians and statisticians have had little to say about factor analysis; mathematicians because the basic procedures have long been known and are relatively trivial, statisticians because the statistical problems of factor analysis such as that of estimating sampling errors, have appeared more or less beyond solution. To statisticians generally, factor analysis has exerted very little appeal; to some it is a dirty word.

The raw material of factor analysis is a table of original measurements. Suppose that we have many individuals a, b, c—to n and that we have measured each of them in various ways A, B, C—to M, all different. We can record the results in a matrix or rectangular table of numbers, like this:

	INDIVIDUALS					
TESTS	a	b	c	d	...	n
A	X_{Aa}	X_{Ab}	X_{Ac}	X_{Ad}		
B	X_{Ba}	X_{Bb}	X_{Bc}			
C						
D						
.						
.						
.						
M	X_{Ma}	X_{Mb}	X_{Mc}	...	...	X_{Mn}

where X_{Bc}, for instance, denotes the measurement B taken on individual c. This, of course, is just a straightforward tabulation of the result of each of the measurements. Since they may be taken on a very large number of people, this is normally a huge matrix.

Factors are Fewer

Now the object of factor analysis is to find certain new composite dimensions, or factors, say α, β, γ, δ etc., fewer in number than the original measurements, uncorrelated with one another and which contain all the information provided by them. Each individual would have a score on each of these factors though it would not be open to direct observation; instead it would have to be inferred from the original measurements. The result of measurement A taken on individual a has been written in the matrix of original measurements as X_{Aa}. Analogously we can write his hypothetical score on the artificial dimension α as $X_{\alpha a}$, on the dimension β as $X_{\beta a}$ and so on. The problem is to find new dimensions such that any individual's score can be reconstituted by adding together some proportion of each of his hypothetical scores. Moreover, these proportions must be the same for every individual.

These proportions obviously have great importance and are appropriately called *factor loadings.* Thus if X_{Aa} is written in terms of factor scores as:

$$X_{Aa} = pX_{\alpha a} + qX_{\beta a} + rX_{\gamma a} + \text{etc.},$$

then we would write X_{Ab} as:

$$X_{Ab} = pX_{\alpha b} + qX_{\beta b} + rX_{\gamma b} + \text{etc.},$$

where p, q and r, the proportions of the factors α, β and γ that are used to reconstitute the measurement A, are known as the factor loadings of measurement A on the factors α, β and γ.

An important point is that factor analysis only attempts to find factors which can be added in this simple way to get back to the original measurements. It is conceivable that the original measurements could be obtained by some more complicated way of putting factor scores together, but this possibility is not encountered in present day factor analysis. What we have described is an ideal which no procedure achieves.

Two Approaches

The attempt to reconstruct original measurements from a smaller number of factors is in practice rarely, if ever, completely successful. Usually it can only be done approximately and this offers the factor analyst a choice. He can either go on extracting factors until he has enough to reconstitute the original measurements almost perfectly, or he can decide at the outset that he will look for those two, three, or however many factors he prefers, which do the best job of recreating the original measurements. These two approaches are really fundamentally different, though the difference is usually ignored. The first approach, that of taking out as many factors as necessary, starts with the observed data and seeks to represent it in another way in terms of factors uncorrelated with one another. The second approach, that of finding the "best" p factors (where p is some number decided upon beforehand) is an attempt to fit a mathematical model to the observed data. The first approach is more accurately described as *component analysis,* and the term *factor analysis* should perhaps be reserved for the second approach. The confusion between these two basic methods has arisen because the arithmetical routines of both are closely similar. No attempt is made in the remainder of this article to unscramble this well established confusion; both will usually be referred to indiscriminately as factor analysis.

Factor analysts start their work by summarizing the information from the matrix of original measurements by calculating the correlation coefficients between each pair of measurements. This is a starting point in picking out that which is common to the various measurements. The correlation coefficients are then written in the form of a matrix like this:

	A	B	C	D	—	—	M
A	r_{AA}	r_{AB}	r_{AC}	r_{AD}	—	—	—
B	r_{BA}	r_{BB}	r_{BC}	r_{BD}	—	—	—
C	r_{CA}	r_{CB}	r_{CC}	r_{CD}	—	—	—
D	r_{DA}	r_{DB}	r_{DC}	r_{DD}	—	—	—
—	—	—	—	—	—	—	—
—	—	—	—	—	—	—	—
M	—	—	—	—	—	—	—

This matrix is obviously square for there are as many rows as there are columns and usually, since there are many less measurements than people, it is a much smaller matrix than that of the original measurements. Another point is that the part of the matrix above the diagonal running from top left to bottom right (called the *principal diagonal*) is reflected on the other side of it because, for instance, the correlation between C and A, r_{CA} is equal to the correlation between A and C, r_{AC}. Another point is that the principal diagonal consists of terms like r_{AA}, r_{BB}, r_{CC} and so on. Now it might be thought that these entries must necessarily all equal 1, for the correlation of any measurement with itself should be perfect. In this case, however, we are dealing with the *hypothetical* correlation which would be found between two actual measurements. Since there are always at least errors in measurement, we might expect such correlations to be high but not necessarily perfect. Just how the spaces in the main diagonal should be filled is an important issue in factor analysis.

Earlier, in discussing the matrix of original measurements, factor analysis was described as an attempt to express each of the measurements in terms of some linear combination of a lesser number of them where the new artificial measurements, or factors, are uncorrelated. If this were possible then we should describe the measurements as *linearly dependent:* if it is impossible we describe the measurements as *linearly independent.* Now if all the M measurements of the matrix of original measurements are linearly independent, we describe that matrix as having *rank* M. If, however, any one of the measurements could always be obtained by a combination of the others, the matrix would be said to have rank (M − 1) at most. In short, the rank of a matrix is the greatest number of linearly independent rows (or columns) to be found in it. Using these terms, we can say that factor analysis aims to express the matrix of original measurements as a matrix of some reduced rank. Since it is easy to show mathematically that the rank of the correlation matrix is always the same as that of the original matrix (if all elements of the principal diagonal are made equal to one) from which it is derived, the same thing can be said of the correlation matrix.

Factor vs. Component Analysis

At this point, however, it is worthwhile to make some distinction between factor analysis and component analysis. The factor analyst is interested in actually reducing the rank of the correlation matrix to the number of factors he thinks is appropriate in his mathematical model. The component analyst, on the other hand, is not interested in reducing the rank of the correlation matrix. He is interested instead in obtaining factors which are not correlated with one another such that the first will explain as much as possible of the original measurements, the second will explain as much as possible of that left unexplained and so on.

Now the factor analyst wishing to reduce the rank of the correlation matrix has one useful tool. It has been pointed out that the elements of the principal diagonal of that matrix are more or less undefined. If appropriate numbers are put in, the rank of the matrix can, under very general conditions, be considerably reduced. A correlation matrix of rank 10, for example, can be reduced to rank 6. That is, we can reduce the results of 10 measurements to scores on 6 factors, whatever the original scores are. In fact the correlations actually found between the different measurements may help to reduce the rank of the matrix still further.

There are, however, two points of view about these diagonal entries, which factor analysts call *communalities.* One point of view expressed by a statistician is:

> Where the [communalities] are completely unknown one method of approach has been to regard them as being at choice; and in particular to assume that they are such as to minimize the number of factors. In general, this seems to assume, on Nature's part, a much more indulgent behavior than we have any right to expect, but it is interesting to see what happens in such cases. (Kendall, 1957, p. 43.)

The other point of view, often expressed by psychologists, is that the communalities are unique numbers which represent that portion of each measurement which correlates with other measurements under consideration; they are not observed, but can be computed. This, however, seems to be a way of describing an aspect of the mathematical model of factor analysis rather than a statement of observable fact.

There is another useful way of looking at the meaning of the rank of a matrix. Suppose, for instance, that we have only two measurements for each of a number of individuals. We could, if we chose, represent the pair of measurements for each person as points on a graph. The distance of the point from one axis could be used to represent one measurement; the distance from the other axis could stand for the other measurement. Points then, represent persons and because their positions depend upon two measurements they are spread over a two dimensional space—a plane. If we had three measurements for each person we would, of course, need three different dimensions in which to represent them, and so on.

To this rule of n dimensions to represent n different measurements there is an interesting exception. If, for example, we have two measurements for each person and the second is, say, always three times the first, then the points representing persons will always fall on one straight line. Because both measurements can be expressed in terms of only one of them, only one dimension is necessary to represent both. Similarly, if three measurements are such that they can be expressed in terms of any two of them, all the points will lie in two dimensions and, in general, n different measurements not linearly independent can be represented in less than n dimensions.

If the original measurements were expressed in terms of factors uncorrelated with one another, we should be able to represent them diagrammatically in a space with as many dimensions as there are factors. In the graph each measurement would be denoted by a point whose distances from the axes (each of which represents a factor) would represent factor loadings. If lines are drawn to join each of these points to the origin the angle between any pair of them can be shown to be a simple function of the correlation coefficient between the measurements they represent (the correlation coefficient is the cosine of the angle). Observed correlations can be represented completely by angles between pairs of lines. In the case of three dimensions we may imagine a series of spokes sticking out in all directions from some fixed point as representing the original measurements—this, of course, is an illustration of the original measurements represented in a three dimensional space, that is by three factors. Now the factors are inferred from the correlation coefficients while the correlation coefficients are derived from observed data. We cannot choose our correlation coefficients; we must put up with those we find. The factors, however, can be chosen in any way that we please, as long as they do the job of representing the original measurements.

Returning to our three dimensional example, the three factors can be represented as three axes perpendicular to one another with their origin at the common point of the lines standing for the original measurements. Now the only observable data expressed in the diagram are the correlation coefficients and these are denoted by the angles between the lines. These angles are, then, the only invariant part of the diagram. But it is clear that if the whole configuration of lines representing the original measurements were rotated with its origin held fixed, the lines would then occupy different positions with respect to the axes and would have to be described differently. In other words, when we represent a correlation matrix in terms of a number of factors, there are an infinite number of ways of doing it. Each of these corresponds to a different rotation of the factor axes relative to the lines representing the original measurements.

Simple Structure

This is a problem which has exercised psychologists greatly. A number of criteria have been suggested for the "best" selection of axes. The most venerated of these is one proposed by Thurstone and called "simple structure." The use of "simple structure" is no more than the factor analytic expression of the principle of parsimony. Essentially, it is that each variable should be represented by as few factors as possible. The rules given by Thurstone for obtaining "simple structure" demand a certain amount of judgment from the analyst. In general "simple structure" means that the axes will be so chosen that a large number of the points representing the original measurements will be near the origin and many of them will be close to one axis or another.

More recently there have been a number of attempts to specify what is meant by "simple structure" in analytical terms. Two well known ones are referred to as the Varimax and Quartimax methods. Roughly speaking the Quartimax method is a method of rotating the axes so that each measurement is described in terms of as few factors as possible. The Varimax method, on the other hand, obtains a rotation of the factor axes so as to minimize the number of measurements in which any one factor occurs.

The various procedures that have been used in factor analysis very often provide different approximations to the same factor solution. Usually they can be translated into one another mathematically or, where their mathematical models differ, their differences are of course, the result of different initial assumptions made by the analyst. In the early days of factor analysis, this was not so clear and disputes between rival analysts were heated and sometimes abusive. One review of rival methods put it this way:

> Factor theory may be defined as a mathematical rationalization. A factor-analyst is an individual with a peculiar obsession regarding the nature of mental ability or personality. By the application of higher mathematics to wishful thinking, he always proves his original fixed idea or compulsion was right or necessary. In the process he usually proves that all other factor-analysts are dangerously insane, and that the only salvation for them is to undergo his own brand of analysis in order that the true essence of their several maladies may be discovered. Since they never submit to this indignity, he classes them all as hopeless cases, and searches about for some branch of mathematics which none of them is likely to have studied in order to prove that their incurability is not only necessary but also sufficient. (Cureton, 1939, p. 287).

Nowadays three procedures seem to share public favor. One of them, cluster analysis, is not a factor analytic method at all. It offers an insight into the way measurements group together in terms of their correlation coefficients. A far simpler procedure than factor analysis it is regarded by many as providing an approximation to it.

Principal Axes Method

The method of factor analysis which has always seemed mathematically preferable is that known as the method of principal axes—this is in fact a component analysis. It achieves a unique resolution of the original measurements into factors. No subjective judgment is called for at any stage of the method. In this method the first axis is selected so that the sum of the squares of the distances of points from the axis is minimized. Successive axes, each perpendicular to the preceding axes, are chosen so as to minimize the squares of the distance of the points from the new axis. Factor loadings on each successive axis become smaller and smaller until the number of factors reaches the rank of the correlation matrix. Usually, the factor loadings on the first few factors can reproduce the correlation matrix very well and the analysis is concluded when such a stage has been reached. Thus the method, a true component analysis, is sometimes used to select a limited number of factors in the same way as a true factor analytic method might.

Centroid Method

The computations required by the method of principal axes are extremely laborious. The most popular method of factor analysis, and one which demands much less labor is the *centroid method*, proposed by Burt (1917) and developed by Thurstone (1931), which was originally intended as an approximation to the principal axes factor solution. Its solution, however, is not unique and some subjective judgment is called for when each set of factor loadings is computed. The *centroid method* selects the first factor axis to pass through the center of gravity of points representing the original measurements. The factor loadings of each measurement on this factor will go a long way toward reproducing the correlation matrix. The difference between the elements of the correlation matrix and those of the attempted reproduction of it, using the first centroid factor, is now regarded as a new correlation matrix and is attacked in the same way. To carry out the next stage, however, it is necessary to alter from positive to negative, or vice versa, all the correlation coefficients involving certain measurements. This selection of measurements for "reflection" of sign, however, is largely subjective.

Currently the principal axes method is gaining favor. Its relative mathematical purity makes it generally desirable and high speed computers have at last made it quite practicable. With the ascendancy of this method the most important questions nowadays in factor analysis are not how it should be done but what it means—and whether it should be done at all.

REFERENCES

BURT, CYRIL. *The Distribution and Relations of Educational Abilities*. London: P. S. King & Son, 1917.

CURETON, E. E. The Principal Compulsions of Factor Analysis. *Harvard Educational Review*, Vol. 9, 1939, pp. 287-295.

HARMON, HARRY F. *Modern Factor Analysis*. Chicago: University of Chicago Press, 1960.

KENDALL, M. G. *A Course in Multivariate Analysis*. London: Charles Griffin, 1957.

RAMOND, CHARLES K. *Factor Analysis in Advertising Research*. Unpublished speech to the Statistical Methods and Operations Research Discussion Group, American Marketing Association, March 30, 1961.

THURSTONE, L. L. Multiple Factor Analysis. *Psychological Review*, Vol. 38, 1931, pp. 406-427.

RESEARCH IN REVIEW

GWYN COLLINS and THORNTON C. LOCKWOOD
ARF Technical Staff

Old Wine in New Bottles

Frank M. Bass et al. (eds.). *Mathematical Models and Methods in Marketing.* Homewood, Illinois: Richard D. Irwin, Inc., 1961. 537 pages, $11.35.

This book consists of 18 papers on the application of mathematical methods to marketing and advertising. All have been published elsewhere and many are already widely known. Here, they are accompanied by editorial commentaries and mathematical appendixes. The scope of the book can be illustrated by a few of the articles it contains: "Variety in Retailing," "A Model for Budgeting Advertising," "An Optimum Geographical Distribution of Publicity Expenditure in a Private Organization," "The Effect of Promotional Effort on Sales." There are other equally interesting articles too, cunningly disguised in the newspeak of operations research.

Since these papers have probably reached almost everyone who can and wishes to understand them, their appearance in this form might seem to call for some explanation. The mathematical appendixes are new, certainly, but these just supply the spade work where the author has offhandedly said, "This expression simplifies to. . . ." They are mere bridge passages in a symphony of symbols.

The editorial commentaries, however, are both new and valuable. They attempt to put into words the assumptions made in the papers themselves. Insofar as they are successful, they expose the whole secret of applied mathematics and operations research.

In every one of these papers the conditions which govern the problem are all common sense ones with which many marketing men would agree. All that has been added by the mathematics is a working out of their consequences. This can be exemplified by one of the best known papers in the collection, "An Operations Research Study of Sales Response to Advertising," by Vidale and Wolfe. This article attempts to describe how advertising and sales are related by the use of three simple concepts.

In the first place they observe that when all advertising is taken away from a product, its sales do not suddenly cease. Instead sales fall off rather rapidly at first, slower thereafter. So they surmise that sales decline in proportion to the volume still left. If this guess is right, then the shape of the sales curve in the absence of advertising is known and can be represented by a mathematical expression which predicts sales level from three facts: the length of time the product has remained unadvertised, its original sales level and some constant which depends on how responsive the product is to the lack of advertising and which can easily be found for individual products.

The second concept is that of an eventual upper limit to the market for the product, a saturation level beyond which sales cannot go.

The third idea is derived from the observation that sales do not immediately reach a stable high level under the impetus of advertising. They increase rather quickly at first, then progressively less and less until some stability is achieved. The authors of the paper assume, rightly or wrongly, that the effect of advertising in increasing sales will be in proportion to the amount by which present sales fall short of saturation level.

Now these assumptions would be rejected by some marketing men. By others they would be hedged about with further conditions and amended here and there. Nevertheless, they are the *kind* of assumptions which a marketing man does make. The mathematics of the paper is then no more than a mechanical working out of the consequences of these assumptions. If these assumptions are indeed correct they lead to conclusions which could not easily be foreseen—and which, unfortunately, are not fully developed in this particular paper. From this paper, for example, we must conclude that a fixed advertising expenditure distributed uniformly over a time period will produce a greater sales return than one concentrated into a single shortlived campaign.

Careful study of these editorial commentaries may do much to convince the non-mathematical reader that operations research is a game that anyone can play and, perhaps, that everyone should. These papers, of course, describe areas where some mathematical thinking has been applied to problems of obvious importance. The authors no doubt wished to show the benefit of such an approach but their case is even stronger than they have made it. For this volume does not, cannot describe all the many occasions when, after a problem has been precisely formulated and numerical values attached to the various desiderata (an activity that has been called "executive psychoanalysis"), its solution is immediately obvious.

It is hard to describe the mathematical level of this book. As the editors admit, it is uneven. Probably the kind of background sometimes known as mathematical economics, at other times as engineering mathematics, would be useful. In any case, a good facility with the calculus will open doors. As for the editorial commentaries, some are delightful expositions—clear, uncomplicated, requiring no mathematics at all. Others are less successful, adding imprecision without removing any difficulties from the papers themselves.

—GWYN COLLINS

Social Science Smörgåsbord

Search. **Published bi-monthly by the National Project in Agricultural Communications. East Lansing: Michigan State University. $5.00 a year.**

"Search . . . for an understanding of human behavior through education and communication" reads the 1961 masthead of this periodical. The National Project in Agricultural Communications began sponsorship of *Search* in 1956, although evidence of anything agricultural seems to have been harvested long since.

Rather, *Search* makes an interesting contribution to social science journalism. In 1960 each issue consisted of four pages, frequently on one topic with a bibliographical reference list, but sometimes containing reviews of four or five books, articles and reports. One issue of the former type analyzed the sources used and the rate of diffusion in the spread of several major news stories. From the 12 studies cited, it was found that news appears to flow largely from the mass media to people without being relayed to any great extent.

Another issue used the results of ARF's *138-Study Summary of the Continuing Study of Newspaper Reading* to conclude that readership patterns, while relatively stable over time, do react to major influences such as the war.

A third issue of *Search* reviewed five recent research writings on a variety of topics: control of speech, social power in organizations, magazine reading, corporate image and group conformity. One review found that conformists and deviants are more difficult to identify in a group whose goals are innovation-oriented. Another reported how verbal cues affect a speaker's choice of words. A third considered how to arbitrate power conflicts in occupational, social and educational groups. A fourth reported on the socio-economic characteristics of readers of general circulation magazines as revealed in the Politz-conducted *Ad Page Exposure in Four Magazines*. The last review reported the ratings of public relations experts on "public images" of corporations compared with the ratings by college seniors of these firms as desirable employers.

In 1961, the format of *Search* was modified as a result of the findings of a subscriber study. The first bi-monthly issue provides a four-page lead-in article by Dr. John T. McNelly of the Michigan State University Communications Research Center. Since the evidence in his article, entitled "Mass Communication and People's Interest," suggests that the mass media do influence behavior and interests, newspapers have a responsibility to present additional information with the news. While this information may provide no immediate reward for the receiver, it may satisfy an information-acquisition drive which the author sees associated with level of education. Upon postulating the receiver's need to synthesize or set new information into a "sensible pattern," the author concludes that the communicator must cloak the information he provides with an immediacy or meaning for the receiver.

Five loose leaf book reviews, three by the editor of *Search*, D. E. Wells, are also enclosed, punched for easy filing. One review considers techniques of associating attitudes with nonsense syllables. Another discusses the influence of administrative conformity on readable writing. A third analyzes the pattern of success ratings among nurses and supervisors. Another review reports different achievement patterns for inbred and imported faculty members. The final study reviewed discusses characteristics of interviewers as determined by supervisory ratings and two personality tests.

Perhaps the most striking aspect of this periodical is the wide scope of its subject interest. There is something in it for students and practitioners of: journalism, broadcast and print media, political science, industrial relations, linguistics, marketing, advertising, sociology and psychology. All such individuals should also be receptive to the more general implications of communication theory and practice.

One can't help wondering if this publication's editorial resources can continue to support such a broad orientation. Many readers will nevertheless be impressed with individual issues that are well written and neatly organized.

—THORNTON C. LOCKWOOD

Correlation Is Not Causation

A. J. Kent and C. Nuttall. An Attempt to Answer the Question—Does Advertising Pay? *The Incorporated Statistician*, Vol. 10, No. 1, February 1960, pp. 29-35.

The authors of this brief paper have tried to express a measure of the sales of two competing British candy products in terms of the current trend of candy consumption, cumulative advertising expenditure on each of the two products, free spending money available, the relative price of the two products, advertising expenditures on all products of the same kind and seasonal effects. After discarding some of these variables as inconvenient or irrelevant, they conclude that advertising and sales are related but that the relationship should be interpreted with great caution, if at all.

An attempt such as this faces four problems. First the researcher must decide what kind of relationship he expects to find; some hypothesis about the way advertising will influence sales is necessary at the outset. Then the very nature of the industry may raise special problems; it may be subject to changing governmental regulations and other such influences which are difficult to treat mathematically. Third, reliable statistics about advertising expenditure, sales and other aspects of a business are not always available. Finally, in interpreting any observed relationship there are always alternative hypotheses which explain the data equally well, questions about cause and effect, and questions of how far the conclusion may be generalized to other products, time periods and situations.

In this paper, the authors hypothesized that advertising expenditures, free spending money, and disproportion between the prices of the two products were all subject to some kind of diminishing returns. In their equation, therefore, the logarithm of all variables dependent upon money were used. They then tried to find linear equations which would predict sales of each product in terms of the other variables.

The British candy industry, the subject of the inquiry, is an awkward one. Strongly competitive, with many products highly substitutable for one another, it has been the victim of some strong outside pressures. From early in the Second World War until 1953, confectionery was rationed in Britain and manufacturers' outputs severely controlled. Then followed a period described by the authors as "post-rationing greed." More recently the industry may have been hurt by the widespread desire to become beautifully less. The authors may, however, have found some consolation in the quality of their data. The Legion Publishing Company has for some years reported British advertising expenditures with accuracy and in detail month by month, and the sales statistics for the two products have, presumably, been obtained from the extensive retail audits of one of the competing companies described.

The results of the study are disappointing, primarily because the relationship which the authors rightly say should be interpreted with caution, hardly emerges at all. Indeed, one can question what interpretation can ever be given to this kind of relationship between sales and advertising. As the authors say, corporation accountants do speak of keeping advertising expenditures "within reasonable proportions," and if sales are not found to relate to advertising expenditures it is certainly not for want of effort by corporation officers with financial responsibility. This kind of study tends to discover company policy rather than the effect of advertising. Nor is it sufficient to try to relate one year's sales to the previous year's advertising, for both are correlated with last year's sales. The solution to this kind of problem lies in forms of deliberate experimentation which have rarely if ever been fully implemented.

The lesson of this paper is, therefore, essentially a negative one. It illustrates the difficulties and futilities of an approach—multiple regression analysis—recently made popular by developments in computing machinery.

—GWYN COLLINS

Missing Link: Attitude—Behavior

Sven Lundstedt (ed.). *Consumer Behavior in 1961, A Summary Report.* Ann Arbor, Michigan: Foundation for Research on Consumer Behavior, 1961. 38 pages, $1.00.

For some 40 years economists have been developing measures of the economy's productive resources and maintaining these measures on a continuing basis. Major improvements in data collecting and processing techniques have permitted the accumulation of these data over shorter time periods. As a result economists have become more and more interested in predicting consumer behavior from one time period to the next. In 1945 the Survey Research Center of the University of Michigan introduced a new time series—measure of buying intentions. With sufficient historical data now accumulated, economists have joined advertisers in asking how attitudes toward buying are related to future purchasing behavior.

This pamphlet reports a one-day conference to explore this relationship, held in January 1961 and conducted by the Foundation for Research on Human Behavior. The conference was attended by 30 individuals, primarily economists and marketing researchers from industry. The title of the pamphlet, emphasizing consumer behavior, appears somewhat at variance with the introduction, which discusses consumer attitudes. The purpose of the conference, however, was to explore the relationship between the two.

In her opening overview, Eva Mueller, program director of the Institute for Social Research, concludes that the deterioration in consumer willingness to buy, as measured by the Index of Consumer Attitudes, has been arrested. Movements of the Index's individual components have tended to balance out. Her discussion is liberally sprinkled with such expressions as confidence, satisfaction, optimism, pessimism, uncertainty, uneasiness, deterioration, caution and alarm, which she uses to describe changes in scale level from one time period to another.

In a second paper, Ruth Mack, economist for the City of New York, demonstrates an interesting relationship between two time series. "All saving," arbitrarily defined, behaves substantially as a constant over time, while a component, durable goods expenditures for automobiles and parts, fluctuates widely. She closes with a plea for basic research on how consumers make up their minds to buy.

From an analysis of consumer stocks of durables —their size, growth and replacement rates—William Butler, vice president of the Chase Manhattan Bank, concludes that a decelerating rate of durable stock replacement and a rising rate of household formation provide evidence for expanding durable goods markets in the next five years. Of the five presentations, his contains the most supporting data.

George Katona, program director of the Institute for Social Research, presents a short review of some of the complexities and perplexities of measuring consumer buying intentions, before coming to the comforting conclusions that in today's society consumer wants are not less insistent, nor the debt burden more onerous, nor the dissatisfaction with available savings greater, than in the last five or ten years.

In the final presentation Vernon G. Lippitt, economist at General Electric, charts a close relationship between consumer expenditures for furniture and household equipment, and the Index of Consumer Attitudes, both by itself and with other relevant series, during 1957-1959. The relationship becomes less close in 1960.

This pamphlet underscores both the distance still to be traveled before being able to predict behavior from consumer attitudes, and the great interest in such travel. While few would dispute the existence of the association, fewer still seem able or willing to predict its magnitude. Nevertheless, measures of consumer buying intentions appear to have a secure position among the various time series used to forecast business activity.

—THORNTON C. LOCKWOOD

Difference in Deference

June Sacher Ehrlich and David Riesman. Age and Authority in the Interview. *Public Opinion Quarterly.* Vol. 25, No. 1, Spring 1961, pp. 39-56.

The June 1961 issue of this *Journal* reported on two seldom discussed sources of possible survey bias, the tendency of some respondents to be overly agreeable, and the tendency of low completion rates by some interviewers to influence over-all survey estimates. The present authors suggest a third source: the age and authority of the interviewer. Any of these sources, if uncontrolled, could introduce bias in a study using competent sampling procedures and questionnaire design.

The data for this analysis were taken primarily from a survey of the attitudes of adolescent girls, conducted in 1956 by the Survey Research Center of the University of Michigan. Using results from four questions with open-end response categories, certain answers were classified as "unacceptable" in the sense that neither the researcher nor the respondent presumably would expect an adult to accept such an answer, nor would the respondent be likely to give such an answer in the presence of a parent or other authority figure.

Two hypotheses were tested:

1. Younger interviewers would obtain more "unacceptable" responses than older interviewers.
2. Younger respondents would give fewer "unacceptable" responses than older respondents.

For purpose of analysis, respondents were divided into three age groups: under 14, 14-15, 16 and over. Interviewers were divided similarly into approximate thirds: 24-40, 41-52, 53 and over. The authors are careful to point out at every opportunity that some cells of their cross-tabulations contain very few cases. They qualify these results accordingly.

Their hypotheses are at least partially confirmed. Specifically, respondents 16 years of age and older gave more "unacceptable" answers to younger than to older interviewers, and respondents 13 years of age and under gave fewer "unacceptable" answers than respondents aged 16 and older.

Answers to three of the four questions used support the above findings. One question, however, shows a reversal. The authors correctly point out one or more of the four questions, taken individually, may be measuring dissimilar dimensions.

Following this major analysis the authors performed *ex post facto* analyses of only those respondents 16 and over. They found, for instance, that younger interviewers tended to have a slightly greater proportion of respondents from traditional homes than did older interviewers. These more traditional respondents in turn produced fewer "unacceptable" responses than did their more modern prototypes, a factor tending to depress differences between older and younger interviewers. On the other hand, younger interviewers tended to elicit slightly more "unacceptable" answers from respondents whose fathers had a college education, whereas older interviewers tended to elicit fewer. Finally, personality test scores for interviewers on the Guilford-Zimmerman Temperament Survey suggest that younger interviewers, having fewer controlling or influencing tendencies (referred to as "nonascendance") and less sensitivity (called "objectivity"), tended to elicit more "unacceptable" responses than did older interviewers who scored higher on ascendance and objectivity.

The authors use this somewhat fragmentary evidence as a basis for stressing the great importance of respondent and interviewer rapport. Matching the ethnic backgrounds of interviewers and respondents has been one method of promoting rapport. Other matching has been done on a social class basis. The authors recommend matching sociopsychological dimensions as well, especially when the subject under surveillance involves semi- or unstructured questioning on a sensitive topic.

Judging from the recent wave of published articles, interviewer selection is drawing considerable critical attention in survey research irrespective of rigorous interviewer training and checking procedures. It seems likely that the rising level of research sophistication may soon make even greater demands on interviewer quality control.

—THORNTON C. LOCKWOOD

Know How Many, But Know Who Too

Harrison B. Summers. Qualitative Information Concerning Audiences of Network Television Programs. *Journal of Broadcasting*, Vol. 5, No. 2, Spring 1961, pp. 147-160.

The first broadcast rating service was introduced in 1929. Despite this relatively short history, many services have come and gone, and others have changed according to the changing needs of the advertising community.

In this informative article Professor Summers reports some of these changes, and emphasizes the trend toward more qualitative program data. In the final portion of his article, he describes qualitative data currently provided by the so-called "big four" rating services—ARB, Nielsen, Pulse and Trendex—and by the National TvQ Service of the Home Testing Institute. His purpose is to document the advertiser's growing need for audience composition characteristics behind the ratings. His premise: advertising man lives not by ratings alone.

As early as 1940, C. E. Hooper was reporting the proportions of men, women and children included in the total program audience. Pulse, ARB and Trendex provided similar information with their first ratings. The author notes that Nielsen alone of the big four was slow to report such breakdowns.

In November 1957, Trendex added to its available audience composition data the proportion of homes in which the respondent could correctly identify the sponsor or sponsors of a program, and whether the program selector was a man, woman or child.

In 1960, Trendex, Pulse and ARB all expanded their periodic reporting of audience composition data. The author points out that this expansion in the quantity of available qualitative audience information reflects keen competition among rating services, as well as a growing user demand for more information about the people programs reach.

This work does not speculate about how ratings or qualitative data are used. One hears that the advertising industry is buying more detailed audience composition data. If so, its increased use might well obviate the specific controversy over ratings.

Madow, Hyman and Jessen, in their report for the House Committee on Interstate and Foreign Commerce, recommended that more audience composition data be furnished regularly rather than on special call. But they recognized that many clients would not use it.

A study of trends in actual usage would provide a valuable complement to this article.

—THORNTON C. LOCKWOOD

BRIEFLY NOTED

W. A. Belson and C. R. Bell. *A Bibliography of Papers Bearing on the Adequacy of Techniques Used in Survey Research.* London: The Market Research Society, 1960. 52 pp.

Good collection of 521 journal articles emphasizing evaluation and analysis rather than description. Entries listed alphabetically by author. Useful subject index emphasizes interviewer and respondent roles, questionnaire design and the mail survey.

Earl A. Buckley. *How to Increase Sales with Letters.* New York: McGraw-Hill Book Company, 1961. 182 pp. $5.00.

This brief but readable recital of standard direct mail procedures features the do's and don't's of different letter approaches together with checkpoint recipes. A good introductory outline.

Nathan Goldfarb. *An Introduction to Longitudinal Statistical Analysis.* Glencoe, Illinois: The Free Press, 1960. 220 pp. $5.00.

Very general discussion of the ways in which panel data are obtained and used. An unilluminating book.

Franklin A. Graybill. *An Introduction to Linear Statistical Models,* Volume I. New York: McGraw-Hill Book Company, 1961. 463 pp. $12.50.

An excellent theoretical reference on experimental design for the researcher of statistical bent prepared to learn some matrix algebra.

Harold Gulliksen and Samuel Messick (Editors). *Psychological Scaling.* New York: John Wiley and Sons, Inc., 1960. 211 pp. $5.00.

A new and important collection of papers presenting recent theory and findings in psychological scaling. Many items in the excellent bibliography will be prerequisite to the study of this book.

Bob R. Holdren. *The Structure of a Retail Market and the Market Behavior of Retail Units.* Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1960. 203 pp. $4.50.

This prize-winning dissertation describes supermarket retailing in terms of economic analysis. It is probably the best study of its kind now available.

W. G. Madow, H. H. Hyman and R. J. Jessen. *Evaluation of Statistical Methods in Obtaining Broadcast Ratings: Report of the Committee on Interstate and Foreign Commerce.* Washington, D. C.: U. S. Government Printing Office, 1961. 163 pp. Free.

A report by three eminent statisticians for Representative Oren Harris's committee. The careful and persistent reader will note, wedged among the misprints, some firm but controversial conclusions. One finds the rating services doing a reasonable job for the money, another that the broadcasting industry needs an Office of Research Methodology to conduct or support comparative studies. There is a 25-page annotated bibliography on TV and radio program ratings and measurements. The ARF plans to publish a digest of this report soon.

Burton Paulu. *British Broadcasting in Transition.* Minneapolis: University of Minnesota Press, 1961. 250 pp. $5.00.

A readable history, description and analysis of the broadcasting industry, both public and commercial, in Great Britain.

M. H. Quenouille. *Rapid Statistical Calculation.* London: Charles Griffin and Company, Ltd., 1959. 250 pp. $5.00.

A useful pocket book which describes quick and easy estimators of variance, correlation and regression coefficients, with a note on the efficiency of each.

Clarence Schettler. *Public Opinion in American Society.* New York: Harper and Brothers, 1960. 534 pp. $7.00.

A discursive account of many aspects of American life illustrating some general concepts in the study of public opinion.

PUBLICATIONS RECEIVED

NAOMI BORETZ
ARF Librarian

The Library has been moved into a larger room at ARF headquarters. We hope our members will visit these new facilities.

The move was made to accommodate our growing collection of marketing research material. This includes an up-to-date subject file of clippings in addition to books and research reports.

The Library is prepared to expand its reference and research service by preparing bibliographies on subjects of interest to members. Members are invited to suggest topics. Thorough literature searches will be done in fields where there is sufficient demand.

Ackoff, Russell. *Progress in Operations Research*, Volume I. New York: John Wiley and Sons, Inc., 1961. 505 pp. $11.50.

Andriot, John L. *Guide to U.S. Government Statistics* (Third Edition). Arlington, Virginia: Documents Index, 1961. 402 pp. $15.00.

Australian Women's Weekly. *The Market for Household Appliances*. AWW Bulletin No. 1. Sydney: Australian Women's Weekly. Free.

Bass, Frank, et al. *Mathematical Models and Methods in Marketing*. Homewood, Illinois: Richard D. Irwin, Inc., 1961. 545 pp. $11.35.

Bharucha-Reid, A. T. *Elements of the Theory of Markov Processes and Their Applications*. New York: McGraw-Hill Publishing Company, Inc., 1960. 468 pp. $11.50.

Brown, Judson Seise. *Motivation of Behavior*. New York: McGraw-Hill Publishing Company, Inc., 1961. 404 pp. $7.50.

Graybill, Franklin A. *An Introduction to Linear Statistical Models*, Volume 1. New York: McGraw-Hill Book Company, 1961. 463 pp. $12.50.

Harman, Harry H. *Modern Factor Analysis*. Chicago: University of Chicago Press, 1960. 469 pp. $10.00.

Kantor, J. R. *An Objective Psychology of Grammar*. Bloomington, Indiana: The Principia Press, Inc., 1952. 344 pp. $6.00.

Luck, David J., Hugh G. Wales and Donald A. Taylor. *Marketing Research* (Second Edition). Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1961. 541 pp. $8.50 Text. $11.35 Trade.

Madow, W. G., H. H. Hyman and R. J. Jessen. *Evaluation of Statistical Methods Used in Obtaining Broadcast Ratings*. Washington, D. C.: U. S. Government Printing Office, 1961. 163 pp. Free.

Marcus-Steiff, Joachim. *Les Etudes de Motivation*. Paris: Hermann, 1961. 164 pp. $2.75.

Mayer, Martin. *The Schools*. New York: Harper and Brothers, 1961. 446 pp. $5.00.

Rubenstein, Albert H. and C. J. Haberstroh. *Some Theories of Organization*. Homewood, Illinois: Richard D. Irwin, Inc., 1961. 492 pp. $10.00.

Schlaifer, Robert. *Introduction to Statistics for Business Decisions*. New York: McGraw-Hill Publishing Company, Inc., 1961. 382 pp. $7.75.

Small Business Administration. *A Survey of University, Business and Economic Research Projects*. Washington, D. C.: Small Business Administration, 1961. 642 pp. $3.50.

Spence, Kenneth. *Behavior Theory and Learning: Selected Papers*. Englewood Cliffs, New Jersey: Prentice-Hall, 1960. 403 pp. $9.00.

Watson, D. and L. Smith. *Operational Research*. O.R. Report No. 13. London: Institute of Practitioners in Advertising, 1961. 31 pp. Free.

Causation in the social science is never simple and single as in physics or biology, but always multiple and complex. It is of course true that one-to-one causation is an artificial affair, only to be unearthed by isolating phenomena from their total background. Nonetheless, this method is the most powerful weapon in the armory of natural science: it disentangles the chaotic field of influence and reduces it to a series of single causes, each of which can then be given due weight when the isolates are put back into their natural interrelatedness, or when they are deliberately combined (as in modern electrical science and its applications) into new complexes unknown in nature. This method of analysis is impossible in social science. Multiple causation here is irreducible.

The problem is a two-fold one. In the first place, the human mind is always looking for single causes for phenomena. The very idea of multiple causation is not only difficult, but definitely anti-pathetic. And secondly, even when the social scientist has overcome this resistance, extreme practical difficulties remain. Somehow he must disentangle the single causes from the multiple field of which they form an inseparable part. And for this a new technique is necessary.

—JULIAN HUXLEY

EDITORIAL

Keep It Crisp

To the extent he is judged by his language, the ad man must be seen as a vague and fanciful fellow. Rightly or wrongly, he is portrayed speaking as one. Some of his euphemisms have become part of American culture: schoolboys know that running it up the flagpole means giving it a try and that success is when everyone salutes.

So it's understandable that euphemism has reared its elsewhere charming head in advertising research. The scientist who survives by persuading ad men soon learns to speak the ad man's language. Well he should. The danger comes when he goes on to use that language in reporting the results of his work.

For language must be appropriate to its user's goal, and the goal of the professional persuader is not the goal of the professional scientist.

Once persuasion is accomplished, the language by which it was done becomes of little moment. The person persuaded acts and the prior sales-talk may be forgotten with impunity by all except historians.

But an experiment or survey is supposed to be a permanent addition to what we know. Knowledge cannot be divorced from the language in which it is reported. Here precise expression is important if for no other reason than that it is something on which others expect to build. Imprecision in scientific language may be as much a sin against posterity as (beware the simile) cheap concrete in a bridge piling.

A more selfish and immediate reason is that research reports are used to make choices which quickly affect the choosers. Few of them would care to sacrifice precision for charm in the information they use. They wouldn't buy a house knowing only that it had scads of room or a stock knowing only that its price was lower than a snake's navel. Yet advertisers buy space and time with the help of presentations in which terms are just as vague.

A common euphemism in advertising research is that which takes the respondent's answer for the fact. We have to keep reminding ourselves that reader means claimed reader and that buyer means claimed buyer. For as Steve Stock puts it on page nine of this issue, we almost never know what the respondent did but only what he says he did. The Internal Revenue Service can attest to the fact that there are differences.

Then when responses are summed and projected, we try to remember that the audience is inevitably the estimated audience because, as Professor Dalenius reminds us on page one, a sample can never represent perfectly any large population. The degree to which it doesn't can be crucial in important media decisions.

But nouns are easy to de-euphemize. Verbs are harder, especially those which imply causality.

Many persons learn in elementary statistics that correlation is not causation. A stimulus-response correlation can mean that S causes R, that R causes S, that S and R are both caused by another factor (or other factors), that S and R varied together by chance—or any combination of the above. Why then do so many correlation-users report that R is determined by, influenced by, affected by, dependent on, or a function of S?

Because these are tempting sentences. They tempt the writing stylist who seeks the specific, the vigorous, the vivid. They tempt the scientist because they stand for his goal: the discovery of repeatable cause-effect links. It takes valor to spurn them for the clumsy but accurate statement that S was associated with, varied together with, or was related to R.

Whenever it can, of course, research should replace discretion as the better part of valor: let the needed experiments be done. Then causal terms may be used with abandon.

Perfect experiments may be fun to report, but they are also expensive. Imperfections in research will always be with us. The trick is to describe them accurately without becoming dull or obscure.

Precision first, then charm where possible. This is the style we have encouraged and tried to use in this *Journal* during its first year. Frankly it's an experiment. You might say we're running it up the flagpole to see who salutes.

—CHARLES K. RAMOND